Development of a custom-designed targeted resequencing gene panel for Parkinson's disease

by

Amokelani Clementine Mahungu

Thesis presented in fulfilment of the requirements for the degree of Master of Science (Human Genetics) in the Faculty of Medicine and Health at Stellenbosch University.

Supervisor: Professor Soraya Bardien

Faculty of Medicine and Health Sciences

Department of Biomedical Sciences

March 2020

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Amokelani Clementine Mahungu

March 2020

Abstract

Parkinson's disease (PD) is a complex neurodegenerative disease characterized by the loss of dopaminergic neurons in a part of the brain known as the substantia nigra. The selective loss of these neurons results in motor impairments also classified as cardinal PD symptoms which are tremor, rigidity, bradykinesia, and postural instability. Additionally, non-motor symptoms also occur namely loss of smell, mood disorders, cognitive decline, sleep disturbances, gastrointestinal and autonomic dysfunction. To date, no cure for PD exists and the underlying pathobiology of the disease is not fully understood. Approximately 90% of PD cases are idiopathic which is proposed to result from a complex interaction of environmental and genetic factors. The remaining 10% of PD cases are defined as monogenic and caused by genes that follow a Mendelian type of inheritance.

Since the discovery of the first monogenic PD-causing gene *SNCA*, numerous genes have been identified and extensively studied in European and Asian populations. These include *ATP13A2, CHCHD2*, *DJ-1*, *DNAJC13*, *DNAJC6*, *EIF4G1*, *FBXO7*, *GBA*, *GCH1*, *GIGYF2*, *HTRA2*, *LRRK2*, *PINK1*, *PLA2G6*, *PRKN*, *RAB39B*, *RIC3*, *SLC6A3*, *SNCA*, *SYNJ1*, *TMEM230*, *VPS13C* and *VPS35*. However, many of these genes have not been widely studied in sub-Saharan African (SSA) populations. Furthermore, for the few that have been investigated, the studies used first-generation sequencing methods namely Sanger sequencing, that only allows screening of a single region or mutation at a time. More recently, next-generation sequencing (NGS) gene panels have been used to examine all the known PD genes in patients whose early disease onset and positive family history suggest a possible inherited genetic cause. Thus, the **first aim** of our study was to design a custom NGS gene panel for rapid screening of the known PD genes in South African patients. The gene panel was developed using Agilent SureSelect Target Enrichment technology and it included all of the above-mentioned 23 PD genes. Subsequently, 32 PD patients with early disease onset and family history were screened in two separate sequencing runs. Following analysis of the sequence data, we achieved a coverage of >200x for both runs. For our first run, we included positive controls with known pathogenic single nucleotide mutations, a 40bp deletion, and copy number mutations. All the mutations from the positive controls were validated except for the copy number mutations. On average, 79 variants were detected per patient of which only 15 were prioritized based on whether they were previously associated with PD, were rare (MAF <0.01), novel, and were predicted to be deleterious by the majority of the *in-silico* tools. These variants were further classified according to the American College Medical Genetics and Genomics (ACMG) recommendations to assess their clinical significance. Only six were found to meet the pathogenic (*GBA* p.L483P [p.L444P]) or likely pathogenic (*GBA* p.R170L [p.R131L], p.D179H [p.D140H], p.E365K [p.E326K], *PINK1* p.P305A, and *PRKN* p.E310D) criteria.

One of the major genetic risk factors for PD and a known cause of Gaucher's disease included on the gene panel was *GBA,* which has a pseudogene (*GBAP1*) whose sequence is about 96% similar. Specialized primers that allow the amplification of only *GBA* are used to avoid the detection of variants in the non-functional

pseudogene. Thus, the **second aim** of our study was to develop a method for screening and validating *GBA* mutations in our laboratory. Once the nested Polymerase Chain Reaction (PCR) and Sanger sequencing method was successfully optimized, 30 of our African Black PD patients were screened. To our knowledge, this is the first *GBA* mutation screening in PD performed within this population. Eight variants were identified of which four were predicted to be deleterious by the majority of the *in-silico* tools. These included three known pathogenic Gaucher's disease-associated mutations (p.R150W [p.R120W], p.R170L [p.R131L], and p.T75del [p.T36del]) of which p.R120W is a known risk factor for PD, and one variant of uncertain significance (p.Q536* [p.Q497*]). In addition, two novel variants (p.F255L [p.F216L] and p.G517R [p.G478R]) were identified of which p.F216L was found to be common (9.9%) in controls. Furthermore, the specialized *GBA* primers were also used to validate all of the prioritized *GBA* variants identified using the gene panel. Only four of the five variants (*GBA* p.L483P, p.R170L, p.D179H, p.E365K) were confirmed to be in the functional gene.

In conclusion, we successfully developed a method for rapid screening of the known PD genes and a technique for screening and validating *GBA* mutations. These methods can be used for rapid and high-throughput screening of the genetic contribution to PD in our local populations and other populations within SSA. Subsequently, these methods will enable us to identify novel candidates for validation in future functional studies. Consequently, this work will also contribute to the development of precision medicine tailored to each PD patient.

Opsoming

Parkinson se siekte (PS) is ’n komplekse neurodegeneratiewe siekte wat gekenmerk word deur die verlies van dopaminergiese neurone in die substantia nigra van die brein. Selektiewe verlies van díe neurone lei tot motoriese afwykings, ook bekend as die kardinale simptome van PS, onder andere bewerigheid, rigiditeit, traer bewegings en posturale instabiliteit. Daarbenewens, het PS ook nie-motoriese simptome soos bui versteurings, kognitiewe afname, slaapafwykings, verlies van reuk en laastens gastrointestinale- en outonomiese versteurings tot gevolg. Tot op hede betaan daar geen kuur vir PS nie en verder bly die onderliggende patobiologie van PS steeds onbekend. Omtrent 90% van PS gevalle is idiopaties en word voorgestel as die gevolg van ’n komplekse interaksie tussen omgewings- en genetiese faktore. Die oorblywende 10% van PS gevalle word as monogenies gedefinieer en volg ‘n Mendeliese oorerwingspatroon.

Sedert die ontdekking van die eerste monogeniese PS-veroorsakende geen, naamlik *SNCA*, is daar al talle gene ontdek en bestudeer in Europese en Asiatiese populasies. Dit sluit in, *ATP13A2, CHCHD2*, *DJ-1*, *DNAJC13*, *DNAJC6*, *EIF4G1*, *FBXO7*, *GBA*, *GCH1*, *GIGYF2*, *HTRA2*, *LRRK2*, *PINK1*, *PLA2G6*, *PRKN*, *RAB39B*, *RIC3*, *SLC6A3*, *SNCA*, *SYNJ1*, *TMEM230*, *VPS13C* en *VPS35*. Baie van die gene is egter nog nie intensief in die sub-Sahariese Afrika (SSA) populasies bestudeer nie. Verder, het studies wat wel al sommige van die gene bestudeer het van eerste generasie volgorde-bepaling tegnologieë, soos Sanger volgorde-bepaling, gebruik gemaak. Dit behels tegnieke wat die sifting van slegs ’n enkele area of mutasie op ’n slag kan handhaaf. In meer onlangse studies word daar van volgende generasie volgorde-bepaling (“NGS-sequencing”) geen panele gebruik gemaak, om al die reeds bekende PS gene in pasiënte met vroeë aanvang en positiewe familie geskiendenis, wat dui op genetiese oorerwing as oorsaak van PS, te ondersoek. Die **eerste doelwit** van ons studie was dus die ontwerp van ’n gespesialiseerde volgende generasie volgorde-bepaling geen paneel vir die spoedige sifting van die reeds bekende PS-veroorsakende gene in die Suid-Afrikaanse populasie. Die geen paneel was ontwerp deur van Agilent SureSelect Target Enrichment tegnologie gebruik te maak en het al 23 van die bogenoemde PS gene bevat. Daarna was ’n siftingstoets, op 32 vroeë aanvang PS pasiënte met ’n familie geskiedenis van PS, in twee aparte volgorde-bepaling lopies gedoen. ’n Dekking van >200x vir beide lopies was bereik na die ontleding van die volgorde-bepaling data. Die eerste lopie het die insluiting van positiewe kontroles met reeds bekende patogeniese enkel-nukleotied mutasies, ’n 40bp delesie en kopie nommer mutasies behels. Al die mutasies van die positiewe kontroles, behalwe die kopie nommer mutasies, was geverifieer. ’n Gemiddeld van 79 variante per pasiënt was opgespoor waarvan slegs 15 geprioritiseer was op grond van vorige assosiasie met PS, skaarsheid (MAF <0.01), nuutheid en of dit deur die meerderheid van die *in-siliko* gereedskap voorspel was om skadelik te wees. Hierdie variante was ook verder volgens die American College Medical Genetics and Genomics (ACMG) aanbevelings geklassifiseer om die kliniesie belang daarvan te evalueer. Slegs ses het aan die kriteria vir patogenies *GBA* (p.L483P [p.L444P]) of waarskynlik patogenies (*GBA* p.R170L [p.R131L], p.D179H [p.D140H], p.E365K [p.E326K], *PINK1* p.P305A, en *PRKN* p.E310D) voldoen.

Een van die hoof genetiese risiko faktore vir PS en ’n reeds bekende oorsaak van Gaucher se siekte wat by die geen panel ingereken was, is *GBA* wat ook ’n pseudogeen naamlik *GBAP1* het, met ’n volgorde oorkeenkoms van 96%. Gespesialiseerde oligonukleotied eksemplare wat lei tot die amplifikasie van slegs *GBA* was dus gebruik om sodoende die opspoor van variante in die nie-funksionele pseudogeen te vermy. Die **tweede doelwit** van ons studie was dus om ’n metode vir die sifting en validasie van *GBA* mutasies in ons laboratorium te ontwerp. Na die suksesvolle optimalisering van die gemodifiseerde polimerasie kettingreaksie- en Sanger volgorde-bepaling metodes, was sifting op 30 van ons swart Afrika PS pasiënte toegepas. Tot ons wete was dit die eerste sifting van *GBA* mutasies in PS wat in die populasie uitgevoer is. Agt variante, waarvan vier deur die meerderheid van *in-siliko* gereedskap as skadelik voorspel was, is geïndentifiseer. Dit sluit in drie reeds bekende patogeniese Gaucher se siekte-geassosieerde mutasies (p.R150W [p.R120W], p.R170L [p.R131L], and p.T75del [p.T36del]), waarvan p.R120W bekend is as ’n risiko faktor vir PS, asook een variant met onbekende beduidenis (p.Q536* [p.Q497*]). Boonop was twee nuwe variant bevindings (p.F255L [p.F216L] and p.G517R [p.G478R]) geïdentifiseer waarvan p.F216L gereeld (9.9%) in die kontroles voorgekom het. Verder was die gespesialiseerde *GBA* oligonukleotied eksemplare ook gebruik om al die geprioritiseerde *GBA* variante, wat deur die geen paneel geïdentifiseer is, te verifieer. Slegs vier van die vyf variante (*GBA* p.L483P, p.R170L, p.D179H, p.E365K) is bevestig om in die funksionele geen voor te kom.

Ter afsluiting, het ons ’n suksesvolle metode vir die spoedige sifting van die reeds bekende PS gene asook ’n tegniek vir die sifting en verifikasie van *GBA* mutasies ontwerp. Hierdie metodes kan gebruik word vir spoedige en hoë-deursettingsvermoë sifting van die genetiese bydrae tot PS in ons plaaslike populasies asook ander populasies in SSA. Die metodes sal ons dus in staat stel om nuwe kandidate te identifiseer wat gebruik kan word vir verifikasie werk in toekomstige funskionele studies. Gevolglik kan die werk ook bydra tot die ontwikkeling van presisie-medikasie wat aangepas word tot elke PS pasiënt.

Acknowledgments

I would like to acknowledge the following individuals who in so many ways contributed to my journey thus far:

To my supervisor, Prof Soraya Bardien, thank you for your invaluable support, guidance, insights, and patience. You are truly an inspiration to us.

To the National Research Foundation (NRF), thank you for your financial support.

To the Tiso foundation, it is such a privilege for me to be part of the Tiso family. Thank you for going above and beyond to ensure that my journey through my masters runs smoothly financially.

To everyone in the division of molecular biology and human genetics and everybody in the magic lab, thank you for your warm smiles and words of encouragement.

To every member of the Parkinson's disease research group, thank you for your support and your willingness to answer any questions I had. A special appreciation to Shameemah, Amica, Minke, Katelyn, Sinead, and Nicola, thank you for your immeasurable kindness, support and in so many ways friendship.

To Ms. Alvera Vorster, thank you for your patience, for guiding me through the library preparation steps, and for your wonderful advice.

To my amazing friends and sisters, I have met here in Cape town, nothing in the world can express how much I appreciate your love and kindness. Thank you for the great adventures, laughs, and stories. My journey here would have been boring and sad without you. To Mary-grace, Josephine, Amogelang, and Zikhona thank you so much for adopting me as a younger sister. I truly appreciate your presence, guidance, support, and friendship.

To my incredible friends in Pretoria, thank you for your words of encouragement and for always believing in me.

To my new but old friend Nthongase Makamo, there isn't enough space to describe how much I appreciate you. Thank you for your presence, for listening, for understanding, for sharing every moment of this journey with me and for always making sure I am doing well. You are truly a gift sent from above.

To my wonderful family, thank you for your encouragement and for believing in me. To my mother, Weldah, thank you for raising me in the best way you knew how. You are my world; I couldn't haven't gotten this far without you. To my grandmothers, thank you for always being the pillar of strength in our family and for always encouraging me to pursue education. To my aunts, uncles, and cousins, thank you for your encouragement, for your prayers, and in so many ways for making sure I have what I need. To my little sisters Nyiko and Wisani, thank you for being the best little sisters in the world. I truly enjoy being your big sister, thank you for the laughs and the stories I get to share with you. A special thank you to my uncles, especially Uncle Carl and Seymour, thank you for your guidance and financial support it has truly helped me so much in my journey.

Lastly, to my heavenly father, without his strength, grace, love, kindness and the joys he has set before me I would have given up. My journey has not been easy but each day, I have found my reason to wake up, to continue, to seek his joy, to find hope, to see the good he has placed in me and others. My faith, hopes, and dreams are solidified in him.

"Trust [confidently] in the Lord forever [He is your fortress, your shield, your banner], For the Lord God, is an everlasting Rock [the Rock of Ages]-Isiah 26:4 (AMP)

Research outputs

Conferences

Poster presentation

South African Genetics Society and the South African Society for Bioinformatics congress 2018

Mahungu, A, Vorster, A, Carr, J, Peeters, A, and Bardien, S. Development of a custom-designed targeted resequencing gene panel for Parkinson's disease. Golden Gates Highlands National Park, Free state, 16-18 October 2018.

Oral Presentation

Southern African Society of Human Genetics congress 2019

Mahungu, AC; Vorster, AA; Bronze, M; Jonathan, C; Gamieldien, J; Bardien, S. Development of a custom-designed targeted resequencing gene panel for Parkinson's disease. Century City, Cape Town, 3-6 August 2019.

Publications

Amokelani C. Mahungu, David G. Anderson, Anastasia C. Rossouw, Riaan van Coller, Jonathan A. Carr, Owen A. Ross, Soraya Bardien., 2019. Mutation screening of the glucocerebrosidase *(GBA)* gene in South Africans of African ancestry with Parkinson's disease. December 2019. Neurobiology of Aging.

Table of contents

List of Figures

List of Tables

List of abbreviations

AAO	Age of onset
ACMG	American College of Medical Genetics and Genomics
AD	Autosomal dominant
ALP	Autophagy-lysosomal pathways
AMP	Association for Molecular Pathology
AR	Autosomal recessive
ATP13A2	ATPase 13A2
BAM	Binary Alignment Map
BWA-MEM	Wheeler Aligner-MEM
BH4	Tetrahydrobiopterin
BLAST	Basic Local Alignment Search Tool
CADD	Combined Annotation Dependent Depletion
CAF	Central Analytical Facilities
CBD	Corticobasal degeneration
CHCHD2	Coiled-coil-helix-coiled-coil-helix domain containing 2
CHRNA7	Neuronal nicotinic acetylcholine receptor subunit alpha 7
CME	Clathrin-mediated endocytosis
CNS	Central nervous systems
CNV	Copy-number variation
COR	Carboxyl-terminal of ROC
DANN	Deleterious Annotation of genetic variants using Neural Networks
DAT-SPECT	Dopamine transporter single-photon emission computed tomography
DBS	Deep brain stimulation
dbNSFP	database for nonsynonymous SNPs' functional predictions
dbSNP	Single Nucleotide Polymorphism Database
DJ-1	Parkinsonism associated deglycase

DNA	Deoxyribonucleic Acid
DNAJC6	Dnaj heat shock protein family member C6
DNAJC13	DnaJ heat shock protein family (Hsp40) member C13
dNTP	Deoxynucleoside triphosphate
DRD	L-Dopa-responsive dystonia
dsDNA	double-stranded DNA
EDTA	Ethylenediaminetetraacetic acid
EIF4G1	Eukaryotic translation initiation factor 4 gamma 1
EtBr	Ethidium bromide
ExAc	Exome Aggregation Consortium
Exo I	Exonuclease I
FATHMM	Functional Analysis Through Hidden Markov Models
FBXO7	F-box protein 7
GBA	Glucocerebrosidase
GBD	Global Burden of Diseases
GCH1	GTP cyclohydrolase 1
GD	Gaucher's Disease
GERP++	Genomic Evolutionary Rate Profiling
GIGYF2	GRB10 interacting GYF protein 2
gnomAD	Genome Aggregation Database
GQS	Genomic Quality Scores
GRB10	Growth factor receptor-bound protein 10
GTP	Guanosine Triphosphate
GTPCH	GTP cyclohydrolase 1 protein
GYF	Glycine-Tyrosine-Phenylalanine
HRM	High-Resolution Melt
HSP40	Heat Shock Protein

HTRA2	Htra serine peptidase 2
IGV	Integrative Genomics Viewer
INAD	Infantile Neuroaxonal Dystrophy
ISP	Ion Sphere Particle
KRS	Kafer-Rakeb Syndrome
LRRK2	Leucine-Rich Repeat Kinase 2
LRT	Likelihood Ratio Test
MAF	Minor Allele Frequency
M-CAP	Mendelian Clinically Applicable Pathogenicity
MetaLR	Meta-Analytic Logistic Regression
MetaSVM	Meta-Analytic Support Vector Machine
MDS	Movement Disorder Society
MLPA	Multiplex Ligation-dependent Probe Amplification
MSA	Multiple System Atrophy
NBIA2	Neurodegeneration with Brain Iron Accumulation Type 2
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
PARK2	Parkin
PARK7	DJ-1
PCR	Polymerase Chain Reaction
PD	Parkinson’s Disease
PDMutDB	Parkinson disease Mutation Database
PET	Positron Emission Tomography
PINK1	PTEN induced putative kinase 1
PLA2G6	Phospholipase A2 group VI
Polyphen-2	Polymorphism Phenotyping v2
PPS	parkinsonian-pyramidal syndrome

PRKN	Parkin
PROVEAN	Protein Variation Effect Analyzer
PSP	progressive supranuclear palsy
PTEN	Phosphatase and Tensin Homolog
QC	Quality Control
RAB39B	Rab Associated Protein
RefSeq	NCBI Reference Sequence Database
RIC3	acetylcholine receptor chaperone
RNA	Ribonucleic Acid
ROC	Ras of Complex Proteins
ROS	Reactive Oxygen Species
SAP	Shrimp Alkaline Phosphatase
SCF	SKP1-Cullin-F-Box
SIFT	Sorts Intolerant From Tolerant
SKP1	S-Phase Kinase-Associated Protein 1
SLC6A3	Solute Carrier Family 6 Member 3
SMS	Sequence Manipulation Suite
SNARE	Soluble N-ethylmaleimide-Sensitive Factor Attachment Protein Receptor
SNCA	Alpha-synuclein
SNP	Single Nucleotide Polymorphism
SNpc	Substantia Nigra Pars Compacta
SNV	Single Nucleotide Variants
SSA	sub-Saharan Africa
SSAHA	Sequence Search and Alignment by Hashing Algorithm
SVS	SNP & Variation Suite
SYNJ1	Synaptojanin 1
TAE	Tris-acetate-EDTA

TMAP	Torrent Mapping Alignment Program
TMEM230	Transmembrane Protein 230
Tris	Tris(Hydroxymethyl)aminomethane
TSS	Torrent Suite software
TVC	Torrent Variant Caller
UKPDSBB	UK Parkinson's Disease Society Brain Bank
UPDRS	Unified Parkinson's Disease Rating Scale
UPS	Ubiquitin-Proteasome System
5'UTR	5' untranslated region
3'UTR	3' untranslated region
Vap	Vascular Parkinsonism
VCF	Variant Call Format
VPS13C	Vacuola Protein Sorting 13 Homolog C
VPS35	Vacuolar Protein Sorting-Associated Protein 35
VUS	Variants of uncertain significance
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
XLID	X-linked Intellectual Disability

Outline of the thesis

This thesis has been laid in four chapters (Introduction, Methods, Results, and Discussion).

Chapter 1 provides an introduction to Parkinson's disease (PD) followed by an extensive literature review of the history, prevalence, diagnosis, treatment, and factors that contribute to the development of the disorder. Furthermore, a brief background on the known genetic factors that contribute to PD is given leading to the aims of the study.

Chapter 2 explains the various methodologies and workflows followed to achieve the outlined aims and objectives.

Chapter 3 is divided into two parts. Part A deals with the development of a custom-designed gene panel. In Part B, the work on screening of the *GBA* gene is provided. This is in the form of a manuscript that was submitted to the journal Neurobiology of Ageing and is currently under review.

Chapter 4 provides a discussion of the results, their implications, and applications, limitations of the study and recommendations for future research.

CHAPTER 1: BACKGROUND

1.1. Introduction

What is Parkinson's disease? Parkinson's disease (PD) is characterized clinically as a movement disorder since the classical cardinal symptoms of tremor, rigidity, bradykinesia, and postural instability are known to impair movement function (Gandhi and Wood, 2005). Additionally, non-motor clinical features also occur, and these include cognitive abnormalities, psychiatric symptoms, gastrointestinal disturbances, and dysautonomia. The pathological hallmarks of PD are the existence of alpha-synuclein aggregates in protein inclusions called Lewy bodies, and the progressive selective loss of dopamine-producing neurons in a region of the brain known as the substantia nigra (Dawson et al., 2003).

It is challenging to decipher the underlying pathways leading to the selective nigral loss of dopamine-producing neurons or build-up of aggregated alpha-synuclein in PD. Pathways that have been implicated in the pathogenesis mainly involve the maintenance of cell homeostasis (Gandhi and Wood, 2005) which has been shown to deteriorate with normal aging (Fedarko, 2011). Mitochondrial function, one of the pathways involved in cell homeostasis, is also found to decline with normal aging. However, in PD patients this decline is accelerated (Winklhofer and Haass, 2010). Thus, mitochondrial dysfunction is postulated to be one of the pathobiological features of PD. Studies have also shown that rapid mitochondrial dysfunction in PD patients results in the accumulation of reactive oxygen species (ROS) (Cui et al., 2012). Other pathways implicated in PD pathogenesis are those involved in clearing accumulated, unwanted or abnormal proteins and cell organelles, such as the ubiquitin-proteasome system (UPS) and autophagy-lysosomal pathways (ALP) (Pan et al., 2008).

About 90% of all PD cases are defined as idiopathic, having no known cause (Deng et al., 2018). It is hypothesized that idiopathic PD results from a combination of environmental and genetic factors that accumulate across time. About 10% of PD cases are attributed to the inheritance of mutations in genes that follow a Mendelian type of inheritance (Lunati et al., 2018). These mutations occur in various frequencies in multiple populations. The first gene associated with the inherited form of PD was detected in 1980 when a mutation was identified in alpha-synuclein *(SNCA)* in a large Italian family (Polymeropoulos et al., 1996). Since then, more genes have been identified as studies focused on recognizing genetic factors contributing to the inherited form of PD (Lunati et al., 2018). Although progress has been made in defining these genetic factors, not all of the PD genes are associated with a clear disease inheritance pattern and complete penetrance. This has added to the complexity of understanding the contribution of these genetic factors in PD pathogenesis and their molecular pathways.

PD prevalence and incidence are on the rise globally. Studies have shown that the global burden of neurodegenerative disorders such as PD has increased significantly as populations are aging. The Global

Burden of Diseases (GBD) 2016 Parkinson's Disease Collaborators performed an extensive systematic analysis of the global prevalence, disability, and mortality rates for PD between 1990 and 2016 (Elbaz et al., 2018). The study reported that worldwide PD prevalence has more than doubled in the past 26 years, from 2.5 million in 1990 to 6.1 million cases in 2016. It is suggested that this upsurge is largely due to the increase in population aging and due to increased life expectancy in PD patients, increased knowledge of the disease and improved access to health care. Although an increase in PD cases and risk is observed across all populations, it is reported to be greater in high-income countries.

More studies on PD are needed in sub-Saharan African populations. Although the prevalence of PD is higher in developed countries, an increase in PD rates in sub-Saharan African (SSA) populations has also occurred. The GBD systematic study recorded an increase of PD prevalence in these populations; the percentage change in age-standardized rates between 1990 and 2016, was 14.3% in southern SSA, 15.9% in western SSA, 21.7% in eastern SSA and 10.1% in central SSA (Elbaz et al., 2018). Aging is a significant risk factor for PD; thus, it is predicted that PD estimates will continue to rise as aging also increases in SSA (Velkoff and Kowal, 2006; Lekoubou et al., 2014), however, few epidemiological studies within this population have been undertaken. Therefore, this suggests that there could be an underestimation of the occurrence of PD in SSA. A review by Williams et al, identified only seven community-based prevalence studies undertaken in SSA populations (Williams et al., 2018). Factors unique to low-income countries such as limited access to health care, low number of specialists, cultural differences and health-seeking behaviour may influence the availability of resources to perform high-quality disease epidemiological studies and consequently the availability of well-defined study participants. This results in high rates of underdiagnosis and misdiagnosis within these populations. However, it is anticipated that as study methods are standardized, and resources are made available to more individuals, more PD cases will be detected. Thus, more studies investigating the underlying aetiology of the disease are urgently needed.

In SSA populations, only eleven studies investigating the genetic basis of PD have been performed thus far (Williams et al., 2018). Furthermore, these studies mostly focused on investigating mutations known to cause PD in patients from European and Asian ancestry. Additionally, not all of the known PD genes were investigated. The genetic architecture of SSA populations is unique and very diverse (Choudhury et al., 2018). Thus, it is essential to not only investigate the known disease-causing mutations but also investigate possible novel mutations in these genes that might be unique to this population. The goal of this study is to address the afore-mentioned knowledge gaps by establishing an NGS method that will enable us to rapidly screen for both novel and known mutations in all the previously identified PD genes. 1.2. Review of literature

1.2.1. The history of PD

Evidence of ancient texts describing Parkinsonism-like features were reported in prehistoric Egyptian, Chinese, Biblical and primeval Indian writings (Manyam, 1990; Zhang, Dong, and Román, 2006; Ovallath and

Deepa, 2013). During the 15th century, in the ancient Indian health system of Ayurveda, PD was described and named Kampavata, which refers to tremor in the classical Indian language of Sanskrit (Manyam, 1990). Kampavata was treated with Mucuna pruriens, or atmagupta (Figure 1.1), which contains about 4-7% of levodopa, a drug now widely used to treat PD.

Figure 1.1 Mucuna Pruriens seeds or Atmagupta in ancient Indian language. The seeds of the plant that contains levodopa, used to treat PD symptoms in ancient Indian medicine and the first-line drug for PD. Its anti-cancer and anti-inflammatory properties have warranted its use in multiple disorders (Lampariello et al., 2012). *Taken from J.M.Garg 2009, licensed by Creative Commons.*

Although PD-like symptoms were previously described in ancient times, the first document on this disorder in Western medicine was written by the late Dr. James Parkinson in the 1800s in his celebrated manuscript 'An essay of the shaking palsy' (Parkinson, 1817). His account of six patients with what he referred to as *paralysis agitans,* contributed significantly to the clinical description of PD. James Parkinson also acknowledged in his essay the earlier accounts of PD by those before him, namely Galen, Juncker, Sylvius de la Boë and Boissier de Sauvage.

PD was only listed as a neurological medical condition by Jean-Martin Charcot (the father of neurology) in 1884, who adapted the name of the disease to honour James Parkinson for his significant contribution to the field (Charcot, 1886). Charcot and his team defined the clinical spectrum of PD and distinguished the disease progression stages marked by tremor and rigidity. Furthermore, they contributed to describing the changes in skeletal structure, the progression of pain and symptoms of the autonomic nervous system in PD.

In 1893, Blocq and Marinesco from Charcot's neurological ward proposed a link between PD pathology and the substantia nigra from an autopsy of a tuberculosis patient with signs of a unilateral tremor (Blocq and

Marinesco, 1893). This was further supported by Brissaud, who studied Blocq and Marinesco's work and also hypothesized that the substantia nigra is destroyed in PD patients (Brissaud, 1895). The presence of protein inclusions called Lewy bodies was firstly described by Freidrich Heinrich Lewy and later confirmed by Tretiakoff who named them after Freidrich Lewy (Lewy, 1912; Tretiakoff, 1919). Additionally, dopamine was found to be a neurotransmitter essential for motor function in PD patients by Arvid Carlsson and colleagues in the late 1950s (Carlsson, 2002).

1.2.2. Prevalence and incidence of PD

The global prevalence and incidence rates of PD have increased over the past decades (Elbaz et al., 2018). It is largely accepted that PD estimates have increased as a result of a rise in population aging. Although numerous PD epidemiological studies have been undertaken across multiple populations thus far, it has proven to be challenging to perform a cross-sectional comparison of PD prevalence and incidence across different ancestral populations (Tysnes and Storstein, 2017). Understanding the epidemiology of PD is necessary as it informs countries on approaches for effective disease burden management.

1.2.2.1 Prevalence

It is widely acknowledged that the prevalence of PD varies between 1 to 2 per 1000 individuals and that it affects 1% of the total population over the age of 60 years (von Campenhausen et al., 2005; de Lau and Breteler, 2006). The prevalence is also estimated to increase by at least 4% by the age of 85 (de Rijk et al., 1995; de Rijk et al., 2000). Recently a systematic analysis by the GBD 2016 Parkinson's Disease Collaborators, reported that the age-standardized prevalence has increased between 2016 and 1990 by 21.7% (Elbaz et al., 2018). Furthermore, the prevalence of PD was found to be higher in high-income countries like North America in comparison to low-income nations like SSA (Figure 1.2). The study also showed that in countries that have undergone rapid industrial growth since 1990 the prevalence of PD has more than doubled. For example, in China (a middle Socio-demographic Index country) that has undergone rapid industrial growth, the percentage change in age-standardized rates was found to be 115.7%, the largest increase worldwide.

1.2.2.2 Incidence

There are few PD incidence studies that have been undertaken to date. Although the reported PD incidence rates vary between 8 to 18 per 100 000 patient-years, there is a lack of well-defined epidemiological characteristics of PD which contributes to the underreporting of incidence rates of the disease (de Lau and Breteler, 2006). The GBD 2016 Parkinson's Disease Collaborators systematic study only identified 34 studies on PD incidence representing nine regions, in comparison to 91 prevalence studies (Elbaz et al., 2018). Some studies hypothesize that the lack of PD incidence studies may be due to case under-ascertainment at older ages because of underdiagnosis, comorbidities, or lack of institutional care (de Rijk et al., 2000).

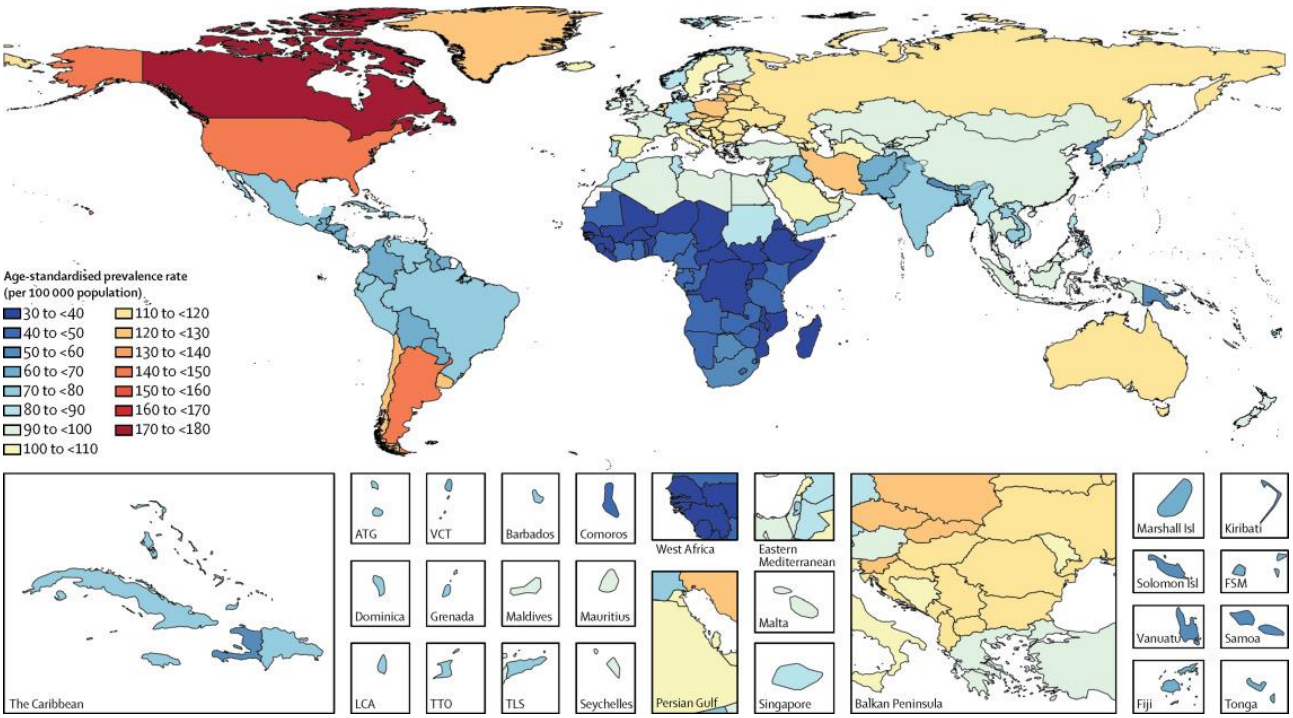


Figure 1.2 Age-normalized prevalence of Parkinson's disease for every 100 000 population. The prevalence of PD is found to be the highest in developed countries such as North America, lower parts of South America, followed by some European countries. The lowest PD rates are reported in West African and some central African countries which form part of SSA. *Taken from GBD 2016 Parkinson's Disease Collaborators licensed by The Creative Commons (https://creativecommons.org/licenses/by/4.0/).*

1.2.2.3 The Epidemiology of PD in Sub-Saharan African populations

SSA consists of African countries that geographically lie south of the Sahara, and countries that makeup SSA are classified as developing or newly industrialized countries. A trend of increased infectious diseases in developing countries compared to a higher prevalence of chronic disorders in high-income or developed nations has been observed over decades (Rocca, 2018). This is because of the economic burden in developing countries resulting in high rates of poverty, pollution, unemployment and limited access to basic health care. The population age structure in developing countries mostly consists of young individuals (Menashe-Oren and Stecklov, 2017). However, studies have shown that life expectancy within these countries has increased and will continue to expand in the upcoming years. It is estimated that by 2050 about 7.6% of SSA populations, which is about 2.074 billion individuals will be above the age of 60 years (Lekoubou et al., 2014). Longer life expectancy and industrialization are positively correlated with increased estimates of chronic disorders, and with the expansion of the populations in SSA and the vast expansion of industrialization, studies predict that the rates of PD will also increase significantly (Williams et al., 2018).

The first epidemiological research studies on PD in SSA populations were conducted in Nigeria, Togo and Ethiopia in the 1980s (Osuntokun et al., 1987; Schoenberg et al., 1988; Tekle-Haimanot et al., 1990; Balogou et al., 2001). These studies were mostly community-based and used a door to door approach. The method was found to be more applicable to SSA populations, as fewer individuals have access to health care systems

(Williams et al., 2018). The prevalence of PD was found to range from 7/100 000 in Ethiopia to 67/100 000 in Nigeria (Williams et al., 2018). Subsequently, two community-based PD prevalence studies were performed post-2000 in Tanzania (Winkler et al., 2010; Dotchin et al., 2008). The prevalence of PD was found to be 33/100 000 by one of the studies (Dotchin et al., 2008). Notwithstanding the significant challenges, more high-quality epidemiological studies in SSA are needed since accurate estimates of PD are essential for accurate disease risk predictions, medical burden projections, and future disease management interventions.

1.2.3. Diagnosis

The diagnosis of PD largely depends on an assessment of the patient's medical history, a clinical physical examination by a specialized movement disorder neurologist and, where possible, the identification of PD-associated lesions from brain scans. PD diagnosis is centered on the occurrence of a combination of PD cardinal motor symptoms, the absence of signs that indicate other diseases and response to levodopa (a drug widely used to treat PD) (Jankovic, 2008). Where possible, radiological imaging of affected brain regions to measure the activity of dopaminergic neurons is performed with dopamine transporter single-photon emission computed tomography (DAT-SPECT) and F-dopa positron emission tomography (PET) scans (Loane and Politis, 2011; Bajaj et al., 2013). The histopathology of post-mortem PD brains is completed by investigating the presence of Lewy bodies in the substantia nigra pars compacta (SNpc) (Gibb and Lees, 1988).

Multiple disease progression rating scales have been developed for characterizing the advancement of PD. The most widely used are the Hoehn and Yahr Scale and the Unified Parkinson's Disease Rating Scale (UPDRS) (Hoehn and Yahr, 1967; Fahn and Elton, 1987). Both scales divide the disease progression into different stages based upon the severity of the signs and have been improved over the years (Goetz et al., 2004; Goetz et al., 2008). The Hoehn and Yahr Scale divides PD progression into five stages, each stage advances as the disease worsens (Hoehn and Yahr, 1967). The UPDRS scale is separated into four parts, each part is individually scored, from zero which indicates normal and 199 indicating severe (Fahn and Elton, 1987). This scale has been updated by Movement Disorder Society (MDS) and it is now referred to as MDS-UPDRS. The MDS-UPDRS added four parts further dividing the scale into a broad category of symptoms (Part 1: Intellectual function, mood, behaviour symptoms; Part 2: symptoms that impair daily activities; Part 3: some motor complications, and Part 4: severe motor complications) (Goetz et al., 2008). Another widely used PD diagnostic criteria was developed by the UK Parkinson's Disease Society Brain Bank (UKPDSBB) and it has been shown to result in a 90% diagnostic accuracy in most cases (Hughes et al., 2001). PD is usually diagnosed when bradykinesia and at least another movement impairment symptom is observed, also considering the lack of signs or history indicative of another diagnosis (Appendix 1).

It should be noted that the diagnoses of PD are challenging because of the presence of some shared phenotypic features with other Parkinsonian syndromes. PD confirmation can thus only be made by validating the occurrence of Lewy bodies in the post-mortem brains of affected individuals. The Parkinsonian

syndromes that share clinical and physiological features of PD include multiple system atrophy (MSA), progressive supranuclear palsy (PSP), vascular Parkinsonism (VaP) and corticobasal degeneration (CBD) (Williams and Litvan, 2013). Misdiagnosis in PD is frequent during the initial consultations and decreases with frequent follow-ups.

1.2.4. Clinical and neuropathological features

1.2.4.1 Clinical features

PD is a clinically heterogeneous disorder; the observed PD features and the affected brain regions may vary among patients. However, there are four cardinal PD motor symptoms: resting tremor, rigidity, bradykinesia and postural instability (Jankovic, 2008). Additionally, freezing gait is also one of the common movement symptoms. Typical PD diagnosis is typically made based on the presence of one or more of these symptoms. The progression and burden of all PD symptoms worsen with age. This results in most affected individuals not being able to continue with their daily activities unassisted.

- Resting tremor

Resting tremor is an involuntary shaking motion and is the first motor symptom that the majority of PD patients report (Hughes et al., 1993). It was also the first symptom James Parkinson highlighted, by naming his essay the shaking palsy which refers to tremor. In most cases, PD is diagnosed once the tremor is observed. A tremor in PD is usually asymmetric, occurring on one side or part of the body and it is most prominent at the distal part of the extremities. It eventually spreads to the arms and legs, depending on where it was initially localized and often occurs when the limbs are at rest. Resting tremor commonly occurs in the hand (Deuschl et al., 2000). While, head tremor generally manifests by shaking of the lips, chin, and jaw (Roze et al., 2006).

- Rigidity

Rigidity is described as an increased resistance to movement (Jankovic, 2008). It is often referred to as a cogwheel phenomenon when it co-occurs with an underlying tremor. PD patients exhibit rigidity by passively flexing, extending, and rotating their bodies. It may occur in the upper body, lower body or the limbs. Rigidity in PD patients may also be accompanied by pain and discomfort.

- Bradykinesia

Bradykinesia in PD manifests as slowed movement resulting in difficulty performing daily activities and reduced reaction times (Berardelli et al., 2001). Impairment of fine motor movement is also observed; patients show slowness in rapid alternating movements such as writing and tying shoelaces. Other manifestations of bradykinesia include hypomimia (reduced facial expressions), difficulties swallowing, speech impairment, and reduced upper limb swing when walking. A couple of patients experience drooling

resulting from difficulties swallowing saliva. Some patients can perform fast movements such as catching a ball when they experience an abrupt flow of emotive energy. Studies suggest that the motor programs in PD-affected brains remain intact; however, patients may have difficulty accessing them. This may also explain how PD patients are able to use previous knowledge to perform an exercise routine, but they cannot initiate or select a movement when not doing the routine.

- Postural instability

Postural Instability appears as a result of the loss of balance and posture control (Jankovic, 2008). Unlike tremor, bradykinesia, and rigidity, postural instability occurs later after the disease progression. Postural abnormalities that result in a bent posture, trunk, and neck, also affect balance and posture. Postural instability, together with freezing (unexpected inability to move) and shuffling gait (slow shuffling movements), contribute to falls and the resulting hip fractures of PD patients (Williams et al., 2006).

- Non-motor features

Studies previously mainly focused on the four cardinal motor symptoms of PD as they are easily visible upon physical examination. However, many have acknowledged the importance of non-motor features in PD diagnosis and progression. These include mood disorders, cognitive decline, impaired gastrointestinal functions, hallucinations, sleep disorders, visual problems, autonomic and olfactory dysfunction (Poewe, 2008). Many PD patients experience cognitive decline, and about 80% of PD cases also report late-onset dementia. The autonomic dysfunctions observed in PD patients comprise of orthostatic hypotension, sweating impairment, erectile dysfunction, and urinary incontinence. Psychiatric disorders reported include depression, anxiety, hallucinations, and apathy. Some PD patients also exhibit abnormal impulsive behavioral symptoms which have been shown to be associated with the dysregulation of dopamine (Gatto and Aldinio, 2019). Other non-movement symptoms reported in PD patients such as loss of smell and constipation seem to occur prior to the onset of movement symptoms and may serve as an indication of the onset of disease progression (Haehner et al., 2007).

1.2.4.2 Neuropathological features

The neuropathological hallmark of PD is characterized by the selective loss of dopamine-producing neurons comprising of neuromelanin in the SNpc (Figure 1.3) and the occurrence of eosinophilic protein inclusions known as Lewy bodies (Brissaud, 1895; Lewy, 1912; Trétiakoff, 1919). Among many pathways, dopamine is predominantly involved in the reward-motivated behavior and control of motor functions (Hornykiewicz, 2008). In PD patients, dopamine and neuromelanin loss results in the depigmentation of the SNpc and the presence of motor symptoms. Studies have shown that the first motor symptoms occur when there is a 60-80% loss of the dopaminergic neurons (Fearnley and Lees, 1991). The progressive neuronal loss in PD also occurs in interconnected brain regions such as the nucleus basalis, locus coeruleus, hypothalamus, cranial

nerve motor nuclei cerebral cortex, and the peripheral and central divisions of the autonomic nervous system.

Although the presence of Lewy bodies in post-mortem PD brains is an essential neuropathological feature for definite PD diagnosis, these protein inclusions are not observed in all PD cases. Lewy bodies are composed primarily of alpha-synuclein, ubiquitin, neurofilament, and alpha B crystallin proteins (Wakabayashi et al., 2013). Lewy neurites, also described in PD pathology, are abnormal neuronal processes comprising of granular material and alpha-synuclein aggregates or filaments (Braak et al., 1999). Alpha-synuclein is the main protein ascribed to PD pathology, and although its function is not completely known, studies have shown that it contributes to PD pathogenesis by forming protein aggregates that spread along interconnected brain regions; and is referred to as the Braak PD staging (Braak et al., 2003). Various studies have indicated that the aggregation of alpha-synuclein may start in the lower parts of the brain stem including the dorsal nucleus of the vagus nerve and the olfactory bulb, progressively spreading to the interconnected brain regions towards the cerebral cortex. Since not all PD patients exhibit Lewy body pathology, some studies do not agree with the Braak hypothesis (Parkkinen et al., 2008). Other studies show that PD is a systemic disease and Lewy body pathology begins at the same time in all affected regions (Engelender and Isacson, 2017). However, cells in different regions of the central nervous systems (CNS) and the enteric nervous system (ENS) may have a variable threshold to the effect of neuronal loss, which in turn corresponds to the manifestations of PD symptoms. For example, in the ENS of the gut, PD symptoms are observed early because low numbers of neurons exist in the gut compared to the SNpc, which requires at least 60% neuronal loss to occur for PD signs to be detected.

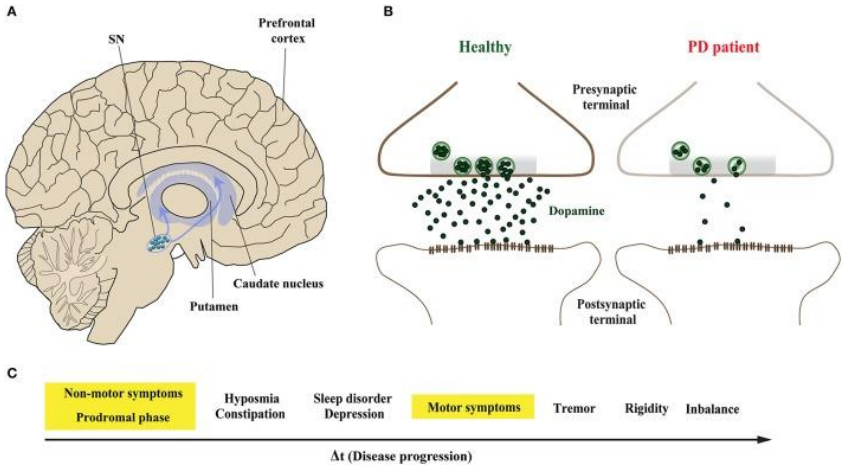


Figure 1.3 The brain region predominantly affected by PD (SNpc) and the difference in the transmission of dopamine in PD patients and non-PD affected persons. A- Shows the substantia nigra and the other brain regions affected by the loss of dopamine. B- Loss of dopamine production and transmission in the synaptic cleft of a PD affected individual *(right)* compared to a non-PD affected neuron *(left)*. C-The selective loss of dopamine transmission progresses with time in PD. This causes the progression of motor symptoms and non-movement related dysfunctions. The non-motor symptoms usually proceed the cardinal motor symptoms. *Taken from* (Bridi and Hirth, 2018) *licensed by The Creative Commons (https://creativecommons.org/licenses/by/4.0/).*

1.2.5. Treatment

No cure currently exists for PD, available therapies or disease management strategies can only alleviate motor symptoms and do not halt neuronal cell loss (Rizek et al., 2016). The treatment and management of PD symptoms include drug medications, surgical procedures, and physical therapy. The widely used and first-line medication for PD is levodopa, a precursor of dopamine (Olanow et al., 2004). Many PD patients respond well to levodopa, however, dyskinesia (impairment of involuntary movements) and dystonia (an uncontrollable contraction of muscles) occur with prolonged use of the drug, especially in patients who present with a more heterogeneous disease phenotype or early-onset PD (Pandey and Srivanitchapoom, 2017; Thanvi et al., 2007). Other drug medications used to treat PD symptoms include dopamine agonists which have been shown to mimic dopamine function (Davie, 2008).

The surgical procedure used to treat PD motor symptoms is deep brain stimulation (DBS) (Figure 1.4), which entails the positioning of one or multiple electrodes into the brain to target certain basal ganglia nuclei (Benabid, 2003; Okun, 2012). DBS is often recommended to PD patients who do not respond well to levodopa. However, treatment is costly and is associated with many risks including bleeding of the cranium (Okun, 2012). Physical therapy used in PD includes coordinated activities or exercise techniques such tai chi, yoga, meditation, massages, and other movement-related physical activities to alleviate shuffling gait, balance disturbances and loosen muscles (Keus et al., 2007).

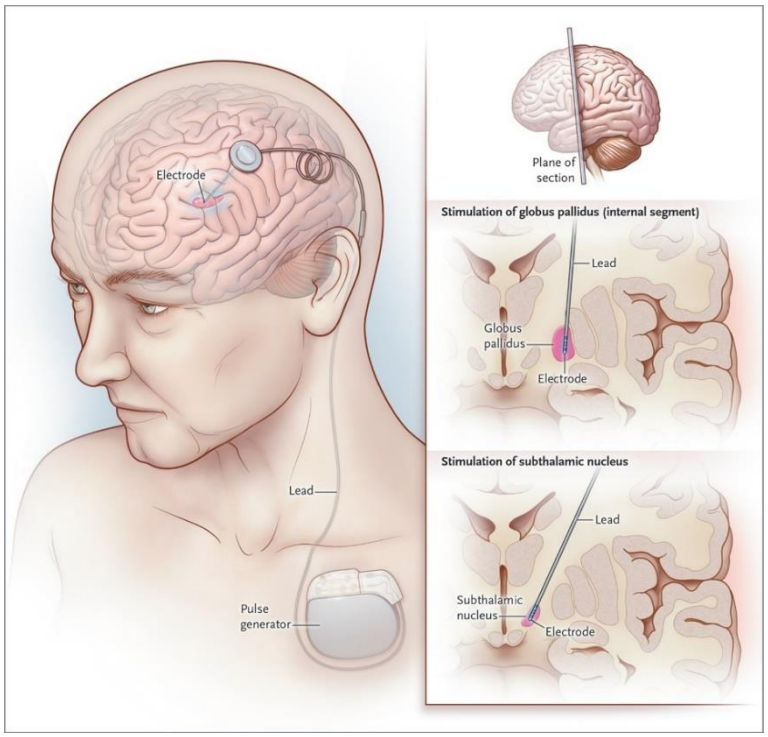


Figure 1.4: Deep brain stimulation. The electrode is implanted through the skull to the subthalamic nucleus. An internal pulse generator is placed under the skin in the top chest of patients to control the stimulation. *Taken with permission from* (Okun, 2012).

1.2.6. Non-genetic risk factors for PD

Although the underlying pathogenesis of PD is still not fully understood, multiple etiological factors have been identified. Studies have suggested that complex interactions of biological (including age and sex) and environmental factors (including neurotoxins) may contribute to the disease risk (Alexander, 2004).

1.2.6.1. Age

Age is shown to be the most common risk factor for developing PD. Thus, the prevalence of PD is suggested to be higher in individuals above the age of 60 years (de Lau and Breteler, 2006). The disease progression is observed to worsen with increasing age (Hindle, 2010). It is unknown how aging increases the risk of PD. However, studies hypothesize that there is an accumulation of age-related damage to cellular functions that may result in disease or reduce the ability of cells to recover from the age-related damage (Reeve et al., 2014). Rare forms of juvenile PD [with age at onset (AAO) ranging between 7 and 21 years] and early-onset PD (< 50 years) have also been reported (Schrag and Schott, 2006). These forms are often linked to a genetic etiology.

1.2.6.2. Sex

The incidence of PD has been found to be higher in males compared to females, it is estimated to be about 1.5 times more common in males (Baldereschi et al., 2000). It has been suggested that the production of estrogen, in females, has a neuroprotective effect. Alternatively, the presence of PD susceptibility genes on the X chromosome of males may increase their PD risk (Inestrosa et al., 1998; Pankratz et al., 2002). Some epidemiological studies suggest that gender roles that affect the occupational environment and lifestyles of both sexes may have resulted in an increased PD risk in men (Tanner and Goldman, 1996). For instance, males mostly work in industries where they are exposed to metals or toxic compounds. However, gender roles are continuing to change over the past decades, and with the increasing exposure to potentially hazardous chemicals in women, it is hypothesized that PD risk may increase in women.

1.2.6.3. Environmental risk factors

In previous studies, it was widely accepted that PD pathogenesis might be largely driven by exposure to environmental pollutants. Previous epidemiological studies have shown that exposure to toxins or heavy metals from the air or water supplies may result in the degeneration of dopamine-producing neurons (Nandipati and Litvan, 2016). This hypothesis was backed by the discovery of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), a toxic chemical accidentally found when a small group of drug users exhibited acute PD-like symptoms as a result of using MPTP contaminated heroin (Langston, 1985). This also led to the development of MPTP-induced PD animal and cellular models, used in research where the selective dopamine loss resembles that of human PD patients (Nandipati and Litvan, 2016).

Previous evidence has also shown that paraquat, a chemical found in pesticides could also result in neurodegeneration and PD pathology. Interestingly, the chemical structure of this compound is found to resemble that of MPTP (Thiruchelvam et al., 2000; Wang et al., 2011). Other environmental risk factors include being exposed to heavy metals mainly iron and manganese, or previous history of head trauma (Nandipati and Litvan, 2016; Factor and Weiner, 1991). Secondary factors, such as exposure to contaminated water supplies and air pollution-producing industries are linked to an increased risk of PD; especially in rural areas because of increased exposure and frequent pesticide use (Nandipati and Litvan, 2016).

1.2.7 Genetic causal and risk factors for PD

Although 90% of all PD cases are idiopathic, 5-10% is attributed to monogenic genes following a Mendelian type of inheritance. It was originally believed that PD is only initiated by exposure to environmental factors, until the discovery of the first PD causing gene, *SNCA*, a gene that encodes the alpha-synuclein protein which is widely found in PD pathology (Polymeropoulos et al., 1997). Since then many more genes, and the susceptibility risk factors, causing monogenic forms of PD have been discovered (Deng et al., 2018; Lunati et al., 2018). Discovery of these genes has shed light on the pathways implicated in PD which includes mitochondrial dysfunction, toxic protein accumulation, oxidative stress and vesicle recycling (Figure 1.5).

The modes of inheritance for monogenic PD genes are autosomal dominant (AD), autosomal recessive (AR) pattern or X-linked. An AD inheritance pattern occurs when a single abnormal copy of the gene causes disease. For an AR inheritance pattern, two copies of the abnormal gene result in PD. An X-linked inheritance pattern occurs when the transmitted mutated gene is found on the X chromosome. The monogenic PD-associated genes that show an AD inheritance pattern are usually associated with late-onset and typical PD symptoms. While the monogenic PD genes showing an AR inheritance pattern mostly occur in PD cases with a juvenile or early disease onset with atypical clinical characteristics such as severe cognitive decline, psychiatric disturbances, or other forms of parkinsonism comorbidities. AR juvenile-onset PD is an acute form of early-onset PD that frequently occurs before the age of 20 years.

So far more than 27 genes are known to cause monogenic forms of PD, however, for the purposes of this thesis and the genes selected for the panel, we will only discuss 23 of these. The other four genes will not be discussed because they are new genes and some have limited evidence of a PD association (Lunati et al., 2018). The ADPD genes highlighted here are, *SNCA*, *LRRK2*, *VPS35*, *GCH1*, *DNAJC13*, *TMEM230*, *HTRA2*, *GIGYF2*, *CHCHD2*, *EIF4G1*, and *RIC3*. Another important PD gene is *RAB39B*, associated with an X-linked disorder with parkinsonism features. The genes associated with ARPD and typical PD features included are *PRKN*, *PINK1*, *DJ-1*, and those associated with atypical parkinsonism are *ATP13A2*, *PLA2G6*, *FBXO7*, *DNAJC6*, *SYNJ1*, *VPS13C,* and *SLC6A3*. Finally, the *GBA* gene, a significant risk factor for PD, will also be discussed. The functions, as well as the predominant clinical features associated with mutations in these genes, will be deliberated.

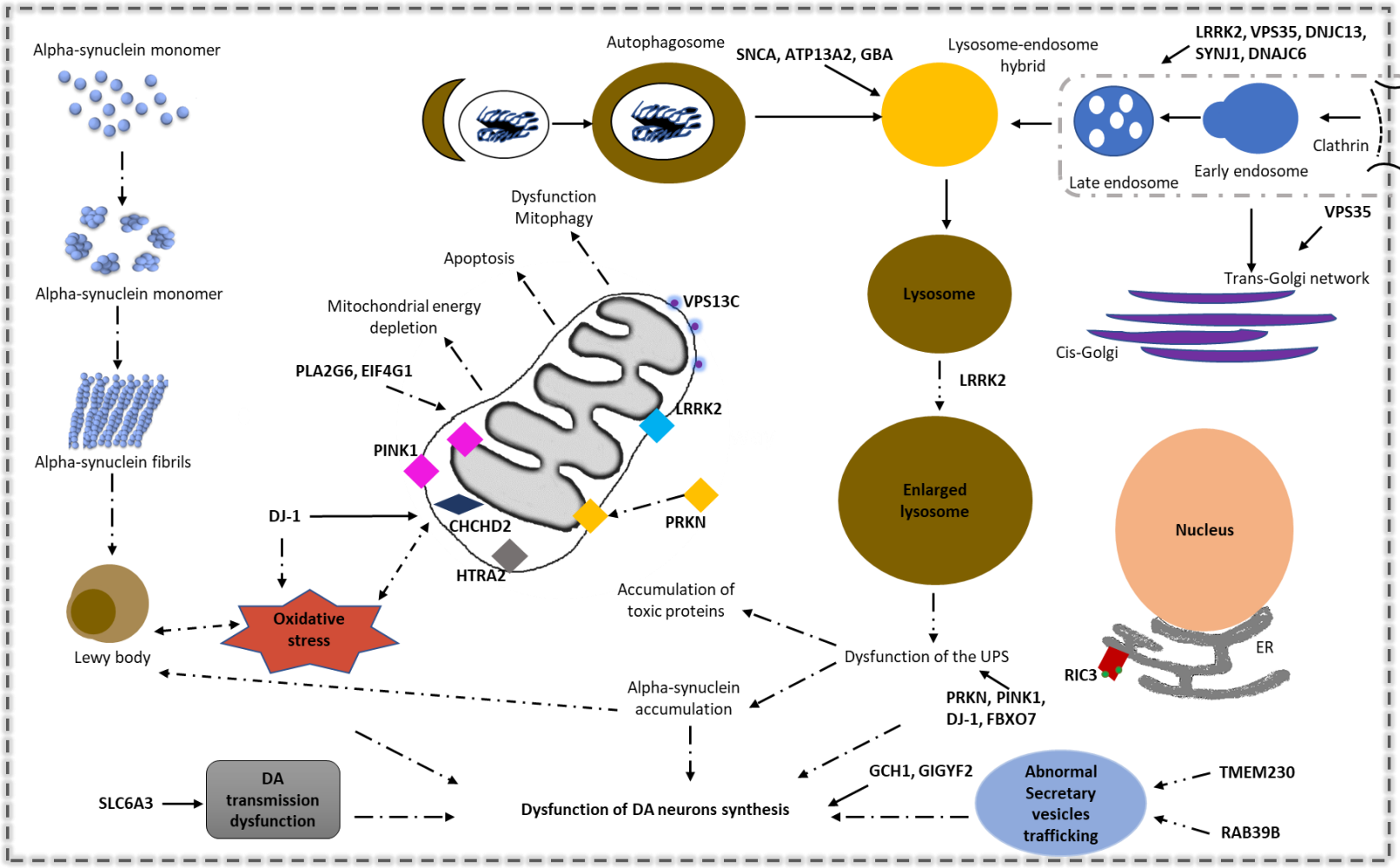


Figure 1.5 Molecular pathways implicated in PD along with the known PD genes. Molecular pathways that result in PD pathogenesis include the maintenance of mitochondrial function, oxidative stress, alpha-synuclein aggregation, as well as the dysfunction of the UPS pathway. The known PD genes have been implicated to play a role in these pathways. The disruption of these pathways is thought to cause loss of dopamine-producing neurons that results in PD symptoms. UPS-Ubiquitin-Proteosome pathway, ER-Endoplasmic Reticulum, DA-Dopamine neurons. The black arrows represent the normal physiological functioning, while the dotted arrows characterize the pathogenic paths. The light grey dotted lines represent endocytosis and the darker dotted lines indicate the cell membrane.

1.2.7.1 Confirmed autosomal dominant PD Genes

- *SNCA* (Alpha-synuclein)

SNCA encodes the alpha-synuclein protein, known to participate in the assembly of the SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) complex and trafficking of the synaptic vesicles (Burre et al., 2010). Although not much is known about the exact function of alpha-synuclein, multiple cellular functions have been assigned to it such as the suppression of apoptosis, regulation of glucose levels, chaperone activity, neuronal differentiation and dopamine biosynthesis (Ahn et al., 2002; Emamzadeh, 2016). Though the normal functioning of alpha-synuclein is not fully elucidated it has been shown that the abnormal protein can be neurotoxic to dopamine neurons by forming aggregates that also result in pathogenic alpha-synuclein monomers and fibrils that are insoluble (Figure 1.5) (Poewe et al., 2017).

The first PD-causing *SNCA* mutation (p.A53T) was detected in a large family of Italian ancestry, and later in five unrelated Greek families with ADPD (Polymeropoulos et al., 1996; Polymeropoulos et al., 1997). Following this discovery, four additional single nucleotide mutations (p.A30P, p.E46K, p.G51D, p.A53T p.A53E), plus duplications and triplications of the whole gene were later identified (Deng and Yuan, 2014; Pasanen et al., 2014; Ferese et al., 2015; Xiong et al., 2016). *SNCA* mutations are associated with a varied disease age of onset ranging from 20-85 years old (Deng et al., 2018). The most common *SNCA* mutation, p.A53T is associated with a younger age of onset (less than 50 years) (Xiong et al., 2016). The p.A30P and p.E46K mutations result in a later onset of PD (age of onset of 60 years) (Krüger et al., 2001; Zarranz et al., 2004). The clinical symptoms observed in *SNCA* mutation carriers also vary (Deng et al., 2018; Lunati et al., 2018). Dementia is dominant in p.A53T mutation carriers, while hallucinations are only observed in p.E46K carriers and severe cerebellar dysfunction symptoms occur in patients carrying the p.A30P mutation. Overall, all *SNCA* PD patients show the cardinal motor symptoms, a positive response to levodopa and similar neuropathology marked by the presence of Lewy bodies and the depletion of neurons in the substantia nigra, and interconnected brain regions.

- *LRRK2* (Leucine-rich repeat kinase 2)

The *LRRK2* gene encodes the leucine-rich repeat kinase 2 protein, which is implicated in multiple functions including the control of autophagy as well as lysosomal pathways (Paisán-Ruíz et al., 2004; Alessi and Sammler, 2018). The *LRRK2* protein is large and is made up of 2527 amino acids. It contains of several domains including the Ras of complex proteins (ROC)-type GTP-binding domain, a carboxyl-terminal of ROC (COR) domain, and the serine or threonine-protein kinase domain (Araki et al., 2018). Studies have suggested that pathogenic mutations in the *LRRK2* gene may result in the loss of dopaminergic neurons by affecting multiple cellular pathways including the lysosomal pathways, protein synthesis, mitochondrial function, and vesicle clearing pathways (Figure 1.5) (Martin et al., 2014).

The *LRRK2* gene locus (12p11.23–q13.11) was initially mapped within a large family of Japanese ancestry exhibiting ADPD (Funayama et al., 2002). Since then, several pathogenic PD causing mutations in the *LRRK2* gene were reported (including p.N1437H, p.R1441G, p.R1441C, p.R1441H, p.Y1699C, p.G2019S, p.I2020T) and their prevalence is found to be population specific. Approximately 1-5% of European ADPD cases are caused by pathogenic *LRRK2* mutations (Lesage et al., 2006; Lesage et al., 2009). The p.G2019S mutation has been observed in over 40% of the North African Arab PD cases (Healy et al., 2008). Although *LRRK2* mutations are generally associated with late-onset PD (> 50 years), age of onset variability also occurs. *LRRK2* mutation carriers have a clinical phenotype often resembling that of idiopathic PD with a good response to levodopa (Thaler et al., 2009; Healy et al., 2008).

- *VPS35* (Vacuolar protein sorting-associated protein 35)

The *VPS35* gene encodes an essential component of the retromer complex, facilitating the regeneration of synaptic vesicles to regulate synaptic endocytosis (Inoshita et al., 2017). Only one *VPS35* mutation, p.D620N is known to cause ADPD (Vilariño-Güell et al., 2011). The mutation was identified in a large Swiss family as well as in Austrian individuals (Wider et al., 2008; Vilariño-Güell et al., 2011; Zimprich et al., 2011). Additional variants in *VPS35* have been identified, but it is not clear whether they are pathogenic (Mohan and Mellick, 2017). The p.D620N mutation is associated with a clinical phenotype comparable to idiopathic PD, with an average age of onset of 50 years (Lunati et al., 2018). The cardinal PD symptoms are initially reported, while the cognitive and psychiatric symptoms are rare. A good response to levodopa in *VPS35* mutation carriers has been observed. A slower disease progression is reported in *VPS35* mutation carriers compared to *SNCA* and *LRRK2* carrying PD cases. It is hypothesized that disruption of *VPS35* functions may result in the interference of the Wnt signalling pathway and the divalent metal transporter 1 which is required to balance iron between endosomes and the cytoplasm (Figure 1.5) (Deng et al., 2013). The Wnt signalling pathway has also been shown to be essential during the growth of dopamine-producing neurons in the midbrain. The disruption of this pathway has been hypothesized to be involved in PD pathogenesis.

- *GCH1* (GTP cyclohydrolase 1)

The *GCH1* gene encodes a cofactor known as GTP cyclohydrolase 1 protein (GTPCH), which is essential for the activity of tetrahydrobiopterin (BH4) required for the synthesis of dopamine (Figure 1.5) (Clot et al., 2009). Mutations in the *GCH1* gene are known to cause childhood-onset L-Dopa-responsive dystonia (DRD). A small dosage of dopamine treatment is often used to treat DRD and a good response is observed. PD has been found to be one of the neurological syndromes occurring in *GCH1* mutation carriers. *GCH1*-related PD is associated with early disease onset and the mean age of onset is 43 years. The observed clinical phenotype includes the presence of the cardinal motor signs and non-motor features such as autonomic dysfunction, cognitive decline, sleep disorders and hyposmia (Mencacci et al., 2014). Neuronal loss and Lewy bodies are also reported in the neuropathological findings of *GCH1* mutation carriers (Gibb et al., 1991).

1.2.7.2 Unconfirmed autosomal dominant PD genes

Numerous genes associated with ADPD that have not been replicated or confirmed have been described and included in screening in PD gene panels (Lunati et al., 2018). It is suggested that these genes may not be common in the widely studied populations. These will be referred to as 'unconfirmed' ADPD genes. Mutations in these genes are also considered a rare causal factor and may occur in different frequencies among diverse populations. Incomplete penetrance of PD is also observed and some of the mutations within these genes have been reported in controls. The unconfirmed ADPD genes include *GIGYF2, HTRA2, EIF4G1,* and newly discovered *RIC3, CHCHD2, DNAJC13, and TMEM230.*

- *GIGYF2* (GRB10 interacting GYF protein 2)

The *GIGYF2* gene encodes the GRB10-interacting GYF protein and it was first mapped using multipoint nonparametric linkage analysis of multiple PD cases with a family history (Pankratz et al., 2003). Later, seven mutations (p.N56S, p.T112A, p.I278V, p.S335T, p.N457T, p.D606E, and p.V1242I) with unknown pathogenicity were detected in 12 unrelated French and Italian PD patients (Lautier et al., 2008). A large meta-analysis of *GIGYF2* studies that involved 5466 PD cases and 6517 controls indicated that it is only a risk factor in Caucasian populations (Zhang et al., 2015). Thus, the involvement of *GIGYF2* in PD pathogenesis remains debated. The GRB10-interacting GYF protein has been shown to be involved in the signalling of insulin receptors and growth factors (Figure 1.5) (Ruiz-Martinez et al., 2015). Moreover, insulin has been suggested to control dopamine neurons.

- *HTRA2* (Htra serine peptidase 2)

The *HTRA2* gene encodes serine protease, found to localize on the mitochondrial intermembrane space (Martins et al., 2004). The gene was initially mapped to chromosome 2p13 and associated with typical idiopathic PD in affected individuals of German ancestry (Gray et al., 2000; Strauss et al., 2005). Later, *HTRA2* mutations were identified by additional studies in individuals with familial and idiopathic PD (Deng et al., 2018). Homozygous mutations in the *HTRA2* gene have also been implicated as a causal factor for AR infantile neurodegenerative disorder, 3-methylglutaconic aciduria (Mandel et al., 2016). The *HTRA2* serine protease has been positioned in the intermembrane space of the mitochondria, coupling with *PINK1* to maintain normal mitochondrial activity (Figure 1.5) (Dagda and Chu, 2009).

- *EIF4G1* (Eukaryotic translation initiation factor 4 gamma 1)

The *EIF4G1* gene encodes the eukaryotic translation initiation factor 4-gamma 1 and it was initially implicated in ADPD in a multi-incident French family that exhibited late-onset PD (Chartier-Harlin et al., 2011). The study also identified novel variants in cases with familial PD and not in controls. However, the identified variants were later also found in healthy controls (Deng et al., 2015). The eukaryotic translation initiation factor 4-gamma 1 regulates the translation initiation of mRNAs that encode mitochondrial genes, are involved in cell

survival and the regulation of growth proteins released in response to stress (Figure 1.5) (Chartier-Harlin et al., 2011).

- *DNAJC13* (DnaJ heat shock protein family (Hsp40) member C13)

The *DNAJC13* gene plays a role in clathrin-mediated endocytosis, stimulating ATP by acting as a co-chaperone of the 'stress-activated' heat-shock proteins (Vilariño-Güell et al., 2014). PD-causing mutations in *DNAJC13* were identified in a large multi-incidental Canadian family of Dutch-German–Russian descent and in six additional families with PD. The *DNAJC13* gene has been suggested to be involved in the endosomal pathway (Figure 1.5) and the mutations in this gene are implicated to alter the transport of endosomes (Vilariño-Güell et al., 2014).

- *TMEM230* (Transmembrane protein 230)

Another study of other members of the multi-incident Dutch-German-Russian family (referred to in the section above) identified mutations in one of the recent PD-associated genes, *TMEM230* (Deng et al., 2016). The *TMEM230* gene encodes a transmembrane protein suggested to play a part in synaptic vesicle trafficking (Figure 1.5). It remains a controversy if both *DNAJC13C* and *TMEM230* segregate with PD in the same family (Deng et al., 2016; Giri et al., 2017). However, mutations in both genes have been identified in a few unrelated PD cases (Appel-Cresswell et al., 2014; Gagliardi et al., 2018; Olszewska et al., 2016; Ma et al., 2017).

- *CHCHD2* (Coiled-coil-helix-coiled-coil-helix domain containing 2)

The *CHCHD2* gene has recently been associated with PD and it encodes a transcription factor that activates and binds to a respiratory chain protein of the mitochondria (Funayama et al., 2015). PD-causing mutations of this gene were identified in a large Japanese family and unrelated PD cases with ADPD. The *CHCHD2* gene has been found to localize in the intermembrane space of the mitochondria (Figure 1.5). Thus, it is hypothesized that it may participate in apoptosis mediated by the mitochondria, oxidative phosphorylation, and neuronal relocation (Funayama et al., 2015; Li et al., 2016).

- *RIC3* (Acetylcholine receptor chaperone)

Lastly, the *RIC3* gene is known to encode a chaperone protein with neuronal nicotinic acetylcholine receptor subunit alpha 7 (CHRNA7) implicated in dopaminergic, glutamatergic and cholinergic pathways that are associated with PD (Figure 1.5) (Sudhaman et al., 2016a). Nevertheless, mutations in *RIC3* were not found in another study of French-Canadian and French PD cases (Ross et al., 2017). Thus, the involvement of *RIC3* in monogenic forms of PD is equivocal.

1.2.7.2 Autosomal recessive genes with typical PD features

- *PRKN* (Parkin)

The *PRKN* gene, also known as *PARK2*, is one of the largest genes in the human genome and it encodes the E3-Ubiquitin ligase protein, named parkin (Shimura et al., 2000). Parkin has a role in the degradation of damaged proteins preserving mitochondrial homeostasis via the proteasome pathway (Figure 1.5). PD-causing mutations in *PRKN* were first identified in unrelated Japanese cases with juvenile-onset ARPD (Kitada et al., 1998). *PRKN* mutations are reported to account for about 77% of all AR juvenile-onset PD cases (Lücking et al., 2000).

To date over 200 *PRKN* mutations have been identified across multiple populations, including missense mutations, deletions, duplications and exon rearrangements (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php). *PRKN* mutations are the largest contributor to early-onset PD (Ferreira and Massano, 2017). They account for approximately 49% of all the early-onset familial PD cases and 18% of young-onset idiopathic PD (Deng et al., 2006). The clinical phenotype of *PRKN*-related PD is characterized by slow disease progression, typical PD motor symptoms with dystonia, a positive response to levodopa, and no psychiatric signs or cognitive impairment (Deng et al., 2006). The neuropathology of *PRKN*-related PD presents a marked loss of neuronal cells within the SNpc and locus coeruleus, however, Lewy body pathology is not commonly found.

- *PINK1* (PTEN induced putative kinase 1)

The *PINK1* gene encodes PTEN-induced putative kinase 1 and is involved in the phosphorylation of parkin to selectively degrade damaged mitochondria (Figure 1.5) (Arena and Valente, 2017; Kazlauskaite and Muqit, 2015; Seirafi et al., 2015). The *PINK1* locus was initially mapped and associated with ARPD in a large Sicilian family (Valente et al., 2001). Later, mutations were reported in three consanguineous Spanish and Italian families (Valente et al., 2004). Thus far, over 100 mutations have been detected in the *PINK1* gene, this includes point, frameshift, and truncating mutations (Arena and Valente, 2017) (HGMD, http://www.hgmd.cf.ac.uk/ac/index.php). *PINK1* mutation carriers exhibit clinical features similar to *PRKN*-related *PD*, with a slightly earlier age of onset. The observed disease phenotype includes a slow progression of the disease, a good response to levodopa accompanied by dyskinesia and dystonia. Additionally, occasional psychiatric signs with minor cognitive decline and Lewy body pathology is observed.

- *DJ-1* (The parkinsonism associated deglycase)

DJ-1 gene also commonly known as *PARK7*, encodes a protein that forms part of the peptidase C65 family found to localize to the cytosol, nucleus and mitochondria (Figure 1.5) (van der Merwe et al., 2015; Wilhelmus et al., 2012). The main function of the *DJ-1* protein is not clear, however, it has been shown to protect cells by regulating oxidative stress. *DJ-1* has also been implicated in the *PINK1*/*Parkin* pathway by transcriptional

regulation of *PINK1* (van der Merwe et al., 2015). *DJ-1* mutations were first detected in a genetically isolated large consanguineous family of Netherlands descent with early-onset ARPD (van Duijn et al., 2001). To date, over 27 *DJ-1* mutations are recorded in the HGMD database (http://www.hgmd.cf.ac.uk/ac/index.php). The mutations include deletions and can occur in a homozygous or compound heterozygous state (Abou-Sleiman et al., 2003; Bonifati et al., 2003). The PD phenotype in *DJ-1* mutation carriers resembles that of *PINK1 and PRKN* related PD cases (Kasten et al., 2018; Kilarski et al., 2012). However, non-movement symptoms such as cognitive decline, psychosis, and mood disorders are more frequent in *DJ-1* carriers. *DJ-1* mutations are the third most common cause of early-onset ARPD after *PRKN* and *PINK1* accounting for about 0.4-1% of all ARPD cases (Kilarski et al., 2012; Abou-Sleiman et al., 2003).

1.2.7.3 Autosomal recessive genes with early-onset atypical parkinsonism

- *ATP13A2* (ATPase 13A2)

The ATP13A2 gene encodes a transmembrane endo-lysosomal-associated protein P5 type transport ATPase (Demirsoy et al., 2017). *ATP13A2* mutations are a known cause of Kafer-Rakeb Syndrome (KRS), a rare neurodegenerative disorder characterized by levodopa-responsive juvenile-onset parkinsonism with dementia, cognitive dysfunction and supranuclear palsy and dystonia (Ramirez et al., 2006). *ATP13A2* mutations are an uncommon cause of juvenile-onset ARPD, and about 23 mutations have been identified in KRS and PD cases. *ATP13A2* symptoms vary among cases with a learning incapacity, kineto-rigid syndrome, impairment of fine motor skills and behavioural disruptions observed in some affected individuals (Lunati et al., 2018). A good response to levodopa is reported, however, long term use results in dyskinesia. The mutated *ATP13A2* gene has been found to localize to the endoplasmic reticulum while the normal gene copy is situated in the lysosome (Figure 1.5) (Demirsoy et al., 2017; Ramirez et al., 2006; Yang and Xu, 2014).

- *PLA2G6* (Phospholipase A2 group VI)

Mutations in the *PLA2G6* gene are known to cause a range of neurological disorders: including infantile neuroaxonal dystrophy (INAD), neurodegeneration with brain iron accumulation type 2 (NBIA2) and autosomal recessive L-dopa responsive parkinsonism with dystonia (Morgan et al., 2006; Gregory et al., 2009). The INAD and NBIA2 disorders occur during infancy (mean age of onset 14 months), while *PLA2G6* related parkinsonism age of onset is usually between 10 and 30 years of age. NBIA2 disorders are characterized by neurodegeneration and the presence of brain iron accumulation. In PD patients with *PLA2G6* mutations, brain iron accumulation is rare. In rare PD cases, brain iron accumulation follows a unique pattern compared to that of NBIA2 cases. The *PLA2G6*-related parkinsonism cases present with a cognitive decline, axial hypotonia, motor regression, spasticity, ophthalmic abnormalities, dysfunction of the bulbar, dystonia and cerebellar atrophy with gliosis (Kurian et al., 2008; Iannello et al., 2017). The *PLA2G6* gene has been implicated in cell homeostasis, cell signalling and mitochondrial function (Figure 1.5) (Kinghorn et al., 2015).

- *FBXO7* (F-box protein 7)

The first *FBXO7* genetic mutations associated with PD were identified in an Iranian family presenting with the parkinsonian-pyramidal syndrome (PPS), a form of parkinsonism with pyramidal signs (dysfunctions of the pyramidal tracts [corticospinal or brainstem neuron fibres that control motor function]) (Shojaee et al., 2008). Later, a few homozygous and compound heterozygous mutations were identified in multiple populations with PPS (Conedera et al., 2016; Di Fonzo et al., 2009; Lohmann et al., 2015). The *FBXO7* protein forms part of the SKP1-cullin-F-box (SCF) interacting with the E3-ubiquitin complex to maintain mitochondrial health by cooperating with *PRKN* and *PINK1* (Figure 1.5). The *FBXO7* mutation carriers present with early-onset severe signs of rigidity, bradykinesia, tremor, pyramidal signs and a positive response to levodopa (Conedera et al., 2016). Similarly to *ATP13A2* mutation carriers, *FBXO7* carriers also experience dyskinesia in response to long term use of levodopa and some have psychiatric signs (Paisán-Ruiz et al., 2010; Yalcin-Cakmakli et al., 2014; Wei et al., 2018).

- *DNAJC6* (Dnaj heat shock protein family member C6)

The *DNAJC6 gene* encodes the putative tyrosine-protein phosphatase, auxilin. Auxilin is involved in clathrin-mediated endocytosis (CME) which is essential for integrating dopaminergic receptors in the cells using clathrin-coated vesicles (Figure 1.5) (Jesús et al., 2014; Köroğlu et al., 2013; Olgiati et al., 2016). A homozygous *DNAJC6* (c.801-2A > G) mutation was first identified in two brothers from a consanguineous family presenting with juvenile-onset ARPD (Edvardson et al., 2012). Later, additional variants were reported in individuals with early-onset PD (Köroğlu et al., 2013; Olgiati et al., 2016). Studies have found that *DNJC6* mutations are an uncommon cause of early-onset PD in specific populations including individuals with Spanish and Chinese ancestry (Jesús et al., 2014; Shi et al., 2016). *DNAJC6* mutation carriers present with an early age of onset, ranging from 7-42 years (Edvardson et al., 2012; Olgiati et al., 2016). Variable typical and atypical PD signs are observed across *DNAJC6* patients (Köroğlu et al., 2013). For example, some experience tremors and bradykinesia, others report late-onset postural instability. Cognitive decline and seizures are described in some individuals, while hallucinations, pyramidal signs, gaze paresis or cerebellar signs also occur in several cases.

- *SYNJ1* (Synaptojanin 1)

SNYJ1 mutations are also considered to be an uncommon cause of early-onset ARPD, with only 12 PD cases from multiple populations reported to date (Puschmann, 2017). The synaptojanin 1 protein has multiple functions including the dephosphorylation of phosphoinositides (phosphorylated inositol lipids) which are essential for cell signalling, membrane trafficking and effective synaptic vesicle endocytosis at nerve terminals (Figure 1.5). *SYNJ1* mutation carriers present with tremors, dystonia, bradykinesia, cognitive decline, dementia, a poor response to levodopa and juvenile-onset seizures (Quadri et al., 2013; Krebs et al.,

2013; Picillo et al., 2014). Overexpression of *SYNJ1* has been associated with mental retardation in Down's syndrome and Alzheimer's disease (Drouet and Lesage, 2014). Additionally, *SYNJ1* mutations are reported in individuals with recessive epilepsy and early-onset severe neurological deterioration (Dyment et al., 2015).

- *VPS13C* (Vacuola protein sorting 13 homolog C)

In 2016, the *VPS13C* gene mutations were reported in three unrelated early-onset PD families (Lesage et al., 2016). All affected individuals presented with a severe disease progression with typical parkinsonism symptoms. Initially, a positive response to levodopa was observed. However, treatment-induced dystonia, pyramidal tracts with brisk reflexes progressing to spastic quadriplegia and a severe cognitive decline later occurred. The vacuolar protein sorting 13 protein localizes on the outer mitochondrial membrane and is implicated in the activity of the mitochondria and vesicle trafficking (Figure 1.5) (Schreglmann and Houlden, 2016). The absence of *VPS13C* activity is associated with reduced mitochondrial membrane potential or mitophagy activity, induced by the *PINK1/PRKN* pathway.

- *SLC6A3* (Solute carrier family 6 member 3)

The *SLC6A3* gene is established to encode DAT, a dopamine transporter (Kurian et al., 2009). The encoded protein has been found to localize to the membrane of some neurons, where it plays a role in transporting dopamine (Figure 1.5). Pathogenic *SLC6A3* mutations are associated with various neuropsychiatric disorders including a dopamine transporter deficiency syndrome, which is a rare disease characterized by movement impairments. The disease mostly occurs during infancy, but rare cases of a childhood-onset have been reported. Pathogenic mutations were also described in two unrelated consanguineous families with infantile parkinsonism-dystonia. *SLC6C3* mutations were also shown to be a major risk factor for PD in a meta-analysis study of individuals with PD (Zhai et al., 2014).

1.2.7.4 X-linked PD genes

- *RAB39B* (Rab associated protein)

The *RAB39B* gene encodes a Rab GTPase which is involved in the regulation of vesicular trafficking in neurons (Figure 1.5) (Wilson et al., 2014). Pathogenic *RAB39B* mutations were first discovered in individuals with X-linked Intellectual Disability (XLID), also presenting with an autism spectrum disorder, epilepsy and macrocephaly. Another form of XLID that presents with early-onset parkinsonism, known as Waisman syndrome was also found to be caused by *RAB39B* mutations. Individuals with Waisman syndrome present with a juvenile-onset intellectual disability and early-onset parkinsonism (Wilson et al., 2014; Puschmann, 2017; Lunati et al., 2018). They experience typical PD symptoms with a good response to levodopa and Lewy body pathology. The *RAB39B* gene is the only X-linked inheritance PD gene and it is still unknown how the abnormal protein results in PD pathogenesis (Lunati et al., 2018).

1.2.8 GBA, a major genetic risk factor for PD

The *GBA* gene encodes beta glucocerebrosidase, a lysosomal enzyme that hydrolyzes glucocerebroside (also known as glucosylceramide) into ceramide and glucose (Holleran et al., 1994). *GBA* mutations that are homozygous and compound heterozygous are a known cause of Gaucher's disease (GD), a recessive lysosomal storage disorder, previously described by Phillipe Gaucher (Gaucher, 1882). Individuals with the disorder show an accumulation of glucocerebrosidase lipid deposits in macrophages (also known as Gaucher's cells). The principal clinical characteristics linked to GD include anaemia, hepatosplenomegaly, thrombocytopenia, bony involvement, and in some cases, neurological signs (Sidransky, 2004).

1.2.8.1 PD risk in GBA mutation carriers

Parkinsonism-like symptoms were initially reported in family members of GD patients (Goker-Alpan et al., 2003). Later, as *GBA*-associated PD gained interest, more cases were identified and investigated. Multiple studies reported a connection between *GBA* mutations and the increased PD risk (Sidransky and Lopez, 2012). The studies also highlighted that PD is more common in heterozygous carriers of the common *GBA* mutations known to cause GD.

1.2.8.2 Clinical features and neuropathology of GBA-related PD

The parkinsonism phenotype in *GBA* mutations carriers is observed to resemble that of idiopathic PD cases (Sidransky and Lopez, 2012). However, PD affected *GBA* carriers typically show an earlier age of onset (around 40 years) (Tayebi et al., 2003). Additionally, non-motor (olfactory dysfunction, depression, anxiety, hallucinations) and cognitive (dementia) symptoms are frequently reported in *GBA* related PD (Riboldi and Di Fonzo, 2019). Post-mortem brain tissue analyses of *GBA*-related PD patients revealed classic PD Lewy body pathology (Velayati et al., 2010). In another study, an alpha-synuclein aggregation pattern following Braak's hypothesis was observed (Kono et al., 2007). It was also reported that PD patients carrying *GBA* mutations may experience a more severe dopaminergic neuronal loss compared to those with idiopathic PD (McNeill et al., 2013).

1.2.8.3 Mutation screening of GBA

The human *GBA* gene is situated on chromosome 1q21, positioned 16kb upstream from its closely related pseudogene, *GBAP1* (Horowitz et al., 1989). The existence of the pseudogene complicates PCR amplification and mutation discovery in *GBA*, as *GBAP1* shares 96% of its exonic sequence with *GBA* (Horowitz et al., 1989; Holleran et al., 1994; Stone et al., 2000). Successful amplification is achieved by designing PCR primers specific to the intronic-exonic regions of the functional gene (Stone et al., 2000). Many previous studies screening for *GBA* mutations in PD patients predominantly focused on identifying the common mutations which occur within exons 8-10 of *GBA*. However, it is recommended that screening of the entire gene would be more informative in other populations in which these common mutations are not frequently encountered (Zhang et al., 2018).

To date, over 300 mutations have been identified spanning the *GBA* gene, including frameshifts, insertions, deletions, splice site variants and recombination alleles classified as severe, null or mildly pathogenic (Hruska et al., 2008). However, only seven (p.N370S, p.R496H, 84insGG, IVS2+1G-A, pV394L, p.L444P, and RecTL) are found to be the most common pathogenic mutations in the Ashkenazi Jewish populations (Gan-Or et al., 2015). Figure 1.6 shows the positions of these seven *GBA* mutations. Interestingly, GD-associated *GBA* polymorphisms (p.E326K and p.T369M) have also been associated with increased PD risk, including in South African PD patients of European ancestry (Velayati et al., 2010; Sidransky and Lopez, 2012; Barkhuizen et al., 2017). The positions of the amino acid that make up the *GBA* enzyme are described using the classical and the Human Genome Variation Society (HGVS) nomenclature system. The classical nomenclature system assigns the 40th amino acid as the first codon, as the first 39 amino acids are cleaved off in the active form of the protein (Hruska et al., 2008). While the HGVS nomenclature counts the first amino acid as the start of the first codon in the protein.

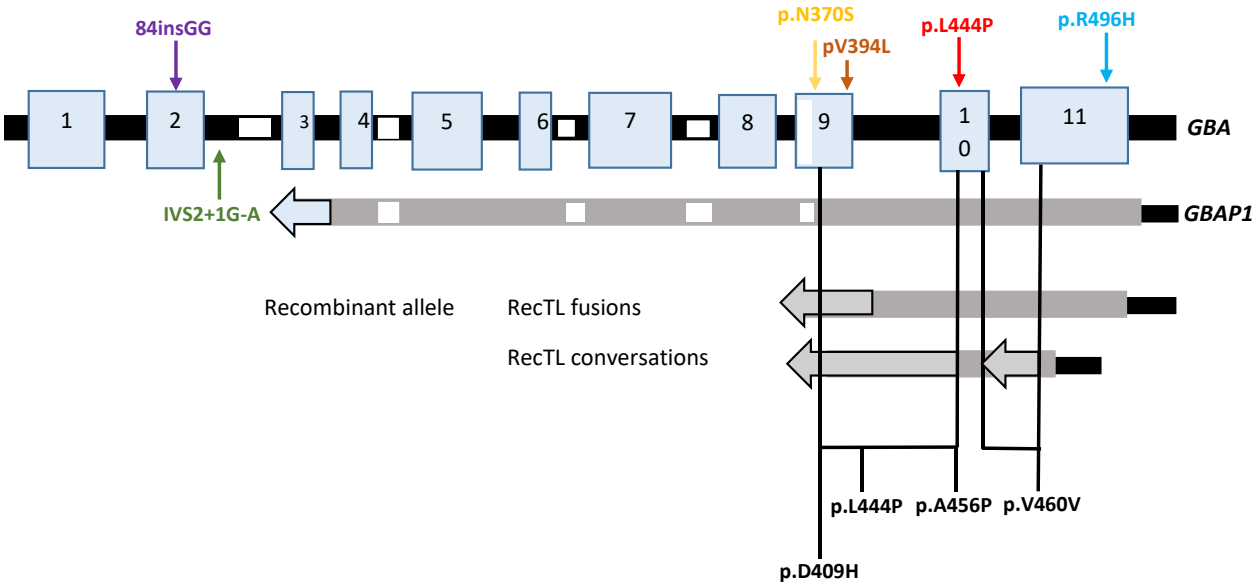


Figure 1.6 The seven-common pathogenic *GBA* mutations in Ashkenazi Jewish populations. The seven most common *GBA* mutations are shown (p.N370S, p.R496H, 84insGG, IVS2+1G-A, pV394L, p.L444P, and RecTL). The splice site mutation in intron 2 (IVS2+1G>A) results in the skipping of exon 2. The recombination alleles occur when there is homologous recombination between *GBA* and the pseudogene and the RecTL recombination allele contains four substitutions (p.L444P, p.A456P, p.V460V, and p.D409H) (Balwani et al., 2011). The sequence of the pseudogene is found to be 96% similar to that of the functional gene and the regions where the two genes do not share homology are indicated by the white boxes. Potential cross over regions on the recombinant alleles is shown by the arrows.

GBA mutations have been shown to frequently occur in Ashkenazi Jewish populations. The prevalence was found to be 19% in Ashkenazi Jewish PD patients screened in one study (Gan-Or et al., 2015). Another study showed that about 18% of this population is a carrier of a *GBA* mutation, while *GBA* mutations occurred in only 1% of the non-Ashkenazi Jewish individuals (Beutler and Gelbart, 1992). Furthermore, the common *GBA* mutations listed above are not frequently occurring in PD individuals of non-Ashkenazi Jewish ancestry (Gan-

Or et al., 2015). Thus, it is hypothesized that unique *GBA* mutations may occur in other populations, highlighting the need for investigating more non-Jewish populations for *GBA* mutations.

1.2.9 Approaches to identify disease-causing genes

The gold standard for mutation screening is Sanger sequencing which was developed and named after Frederick Sanger (Sanger et al., 1977). Sanger sequencing is a first-generation sequencing technology, based on the assimilation of chain-terminating deoxynucleotides triphosphates where the DNA fragments are separated by electrophoresis. Although Sanger sequencing is the benchmark method used for identifying genetic variation, it has scalability limitations. NGS is one of the recent sequencing technologies developed and it has been shown to overcome Sanger sequencing limitations by allowing high-throughput massively parallel sequencing of multiple samples at a lower cost (Mardis, 2011). For our study, we choose this approach as it will enable the investigation of multiple PD genes, in multiple samples in a single run.

The NGS workflow involves multiple steps including sample preparation, template sequencing, alignment of sequence data to reference genome, data filtering and prioritization of interesting mutations (Figure1.7) (van Dijk et al., 2014). The type of NGS workflow followed depends on the sequencing platform employed. To date various sequencing platforms have been developed, the commonly used include the Ion Torrent and Illumina platforms. Ion Torrent sequencing utilizes a semiconductor technology, designed to identify protons released as nucleotides are added during the sequencing while Illumina uses the sequencing by synthesis approach by means of reversible terminator chemistry (Rothberg et al., 2011; Bentley et al., 2008). Although Illumina is more commonly used than Ion Torrent sequencing, both sequencing technologies produce comparatively good quality sequencing data.

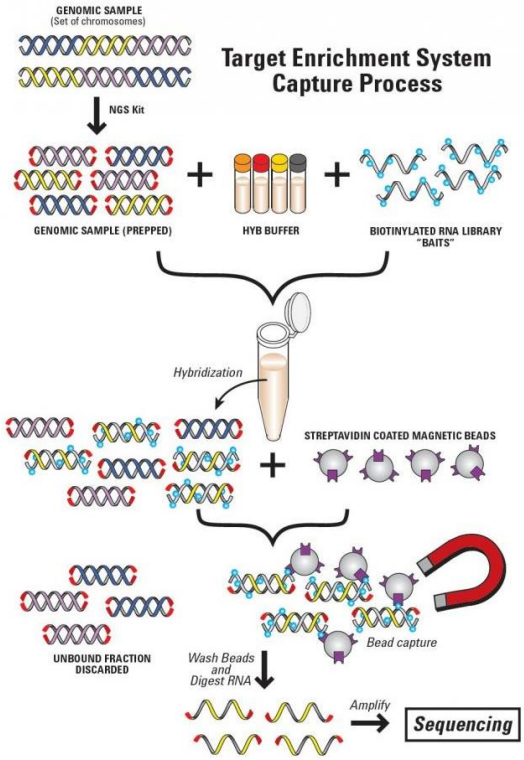


Figure 1.7 Schematic diagram of the basic Target enrichment NGS workflow. The NGS workflow starts with the preparation of samples for sequencing, which includes DNA fragmentation, hybridization, and capture of target regions. This is followed by sequencing on an NGS platform and data analysis**.** *Used with permission from Agilent.*

There are various NGS approaches to screen human genomic DNA and this includes whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted sequencing (Mardis, 2008). WGS focuses on sequencing the complete genome (includes all chromosomal DNA in an individual) while WES involves sequencing of only the protein-coding exons (Bentley, 2006; Wang et al., 2013). While, targeted sequencing focuses on sequencing only regions of interest in the genome (Dilliott et al., 2018). Although WGS and WES produce huge data sets, the specificity of targeted sequencing allows us to only investigate genes of interest. It also allows for the elimination of unrelated genetic variation that can complicate the downstream data analysis (El-Metwally et al., 2013). An increase in coverage also occurs with targeted sequencing. This is essential for the correct identification of variants. For variants that occur at low frequencies having higher coverage increases the ability to accurately detect false positives and it improves sensitivity. Thus, increasing the chances of identifying low frequent variants as real variants. Furthermore, the large datasets produced from WGS and WEG require intensive computational bioinformatics and storage approaches that may be costly. Thus, for our study, we utilized the targeted sequencing approach as it allowed us to only focus on relevant regions of interests that may have clinical significance and limit the complications of sifting through large amounts of data (Figure 1.8).

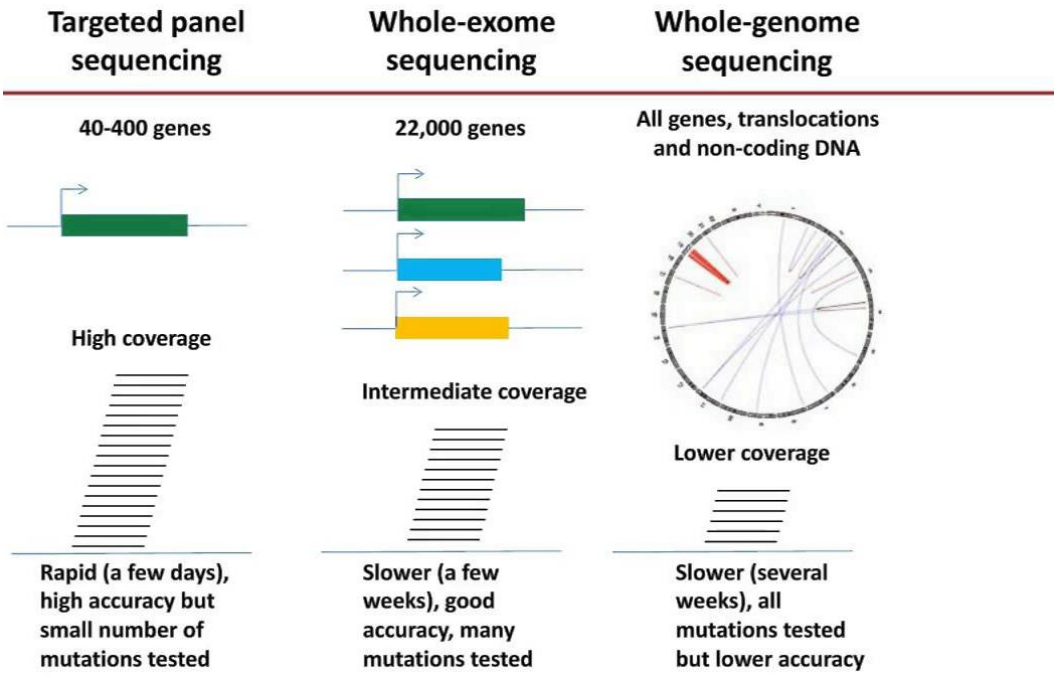


Figure 1.8 A comparison of Targeted sequencing, Whole-exome sequencing (WES), and Whole-genome sequencing (WGS). Targeted sequencing involves the amplification of the regions of interests in the genome, while WES includes amplifying all the coding genes and WGS entails sequencing the whole genome (including non-coding regions). These methods are more advanced than 1st generation sequencing technologies such as Sanger sequencing in that they allow rapid screening of multiple regions for numerous samples simultaneously and produce large amounts of data. Targeted gene panel sequencing is set to be more advantageous than WES and WEG, as it is more rapid, cost-effective, produces less data of higher coverage that is easy to manage and interpret (El-Metwally et al., 2013). *Taken from (Rennert, 2016).*

1.3 The present study

It is anticipated that the identification of the causal genetic factors for PD will aid in establishing a better understanding of both monogenic and idiopathic forms of the disorder. This is essential for the design of therapeutic targets that can be tailored for each patient based on their specific genotype and phenotype. As genetic testing becomes more standardized and sequencing of genomes becomes easier, each individual's genomic, transcriptomic, metabolomic and proteomic data will be utilized to design precision medicine treatment regimens (Mohan and Radha, 2019).

However, precision medicine is far from becoming a reality for PD in SSA populations as the genetic causes of this disorder are poorly characterized in these populations due to limited genetic and epidemiological PD studies (Williams et al., 2018). To date, only 11 PD genetic studies have been conducted in a few SSA countries including South Africa and Nigeria. Of all the known PD-causing genes, only *LRRK2*, *PRKN*, *SNCA, PINK1, DJ-1, ATP13A2, EIF4G1,* and *VPS35* were screened for the known PD-causing mutations. Furthermore, methods that are not high throughput such as Sanger sequencing, Multiplex Ligation-dependent Probe Amplification (MLPA) and High-Resolution Melt (HRM) were utilized.

Given the unique ancestry of SSA, it is hypothesized that novel mutations within the known PD genes may contribute to PD pathogenesis. The novel genetic factors may have been missed by not employing a method that allows screening of all coding regions in all well-established PD genes. Thus, an NGS approach that allows multiplex sequencing of multiple samples and genes is essential. Furthermore, establishing the occurrence of *GBA* mutations in SSA populations of African ethnicity is also essential for better risk estimates and disease management of *GBA*-related PD. To date, only one study has reported on the prevalence of *GBA* mutations in South African European populations (Barkhuizen et al., 2017). To our knowledge, no study has reported on the occurrence of *GBA* mutations in individuals of African ancestry or elucidated on the *GBA-PD* risk within these populations. Investigating *GBA* mutations in SSA populations is essential as *GBA* is a major risk factor for PD and the majority of the previously screened mutations are not found to be common in SSA populations. Furthermore, as therapies that compensate for the lack of *GBA* enzymes are developed, it is anticipated that they will aid in alleviating *GBA-PD* related dysfunctions (Ryan et al., 2019).

The Parkinson's Disease Research Group was established at Stellenbosch University in 2007 with a goal to study the genetic aetiology of PD in South African patients (www.sun.ac.za/parkinsons). Of the 11 genetics studies on SSA populations reported in Williams et al. , the majority (eight) are from this group. To date, this group has recruited a total of 1147 study participants for PD research. This includes 680 PD probands, 43 affected and 383 unaffected family members, and 822 non-PD controls. The present study forms part of this larger parent project, and the aims and objectives of the study are outlined below:

1.3.1 Aims

The aims are two-fold:

1. To establish a custom-designed NGS gene panel to facilitate rapid screening of the well-known PD genes in local South African populations.
2. To set up the methodology for mutation screening of the *GBA* gene in our laboratory.

1.3.2. Hypothesis

We hypothesize that, given the unique ancestry of the South African population, novel pathogenic mutations in the established PD genes will be identified in our patients.

1.3.3 Objectives

For Aim 1:

- Perform a literature search to identify the best candidate PD genes for the panel.
- Select appropriate probands from our collection of PD patients for screening with the panel.
- Set up a custom-designed targeted resequencing panel using SureSelect Custom Target Enrichment technology.
- Use *in-silico* tools to prioritise sequence variants for further study.
- Use Sanger sequencing to validate selected sequence variants.

For Aim 2:

- Set up a nested PCR method for screening of the *GBA* gene.
- Select appropriate PD probands of African ancestry for screening with this method.
- Use various *in-silico* tools to assess the pathogenicity of variants identified.
- Screen ethnic-matched controls to determine the frequency of novel variants found.

CHAPTER 2: MATERIALS AND METHODS

2.1 Ethics approval

Ethical approval for the study was derived from the Health Research Ethics Committee of Stellenbosch University, Cape Town, South Africa (HREC 2002/C059); Appendix 2. Furthermore, written informed consent was obtained from all study participants (cases and controls).

2.2 Selection of study participants

The Parkinson's Disease Research Group at Stellenbosch University has recruited PD patients for genetic analysis since 2007. In addition, family members of these patients (where possible) and ethnic-matched controls have been recruited. The PD patients have been recruited from the Movement Disorders Clinic at Tygerberg Hospital in Cape Town and also from various neurology clinics around the country. All PD affected individuals were assessed by a neurologist and fulfilled the UKPDBBS diagnostic criteria for PD diagnosis (Gibb and Lees, 1988). The controls were recruited from various Western Province Blood Transfusion blood collections clinics around the Cape Town metropole region and these were used to assess the frequency of prioritized variants.

For the part of the study that involved the NGS gene panel, we used a total of 24 PD probands. These were selected based on the following criteria: early age at onset of PD (< 50 years) and/or a positive family history of PD. This was done as these individuals are more likely to have a genetic cause for their disease. Additionally, patients in whom we had previously identified pathogenic mutations were included in the runs as positive controls, as well as non-PD controls as negative controls. Information on the 24 probands, as well as the positive and negative controls, are shown in Appendix 3. The pedigrees of the 24 PD patients are also shown in Appendix 3.

For the second part of our study that involved mutation screening of the *GBA* gene, we selected 30 PD patients of African ancestry. Information on these individuals is shown in Appendix 3.

2.3 DNA extraction

Prior to the present study, DNA had been extracted and made available for use. The DNA was extracted from peripheral whole blood samples using either an in-house method (phenol/chloroform method) or the Nucleo Spin Blood XL kit (Machery-Nagel Duren, Germany) according to the manufacturer's instructions (Appendix 4). Our laboratory had changed over from the in-house method to the commercially available kit method in 2013.

2.4 Candidate gene selection and gene panel design

To initiate the design of the panel, a review of the literature was performed to identify suitable candidate genes. We selected genes in which pathogenic mutations had been identified in more than one affected family member and were not present in controls. Additionally, the families in which these mutations had been identified had to exhibit clear Mendelian inheritance patterns of PD.

The targeted custom gene panel was designed using the Agilent SureDesign software package (Agilent, 2018). SureDesign is a web-based application used to design oligonucleotide probes that capture and enrich the regions of interest, creating the target enrichment library. The probes are manufactured using Agilent's SurePrint oligo manufacturing technology. For our study, we selected the hybrid capture-based target enrichment SureSelect custom panel as it has been indicated to perform better than the amplicon-based panel (Samorodnitsky et al., 2015). The selected SureSelect custom panel was in tier one (library size 1-499 kb). It included SureSelectXT kits which consist of reagents and bait library (probe library) for the post-capture indexing of 16 samples per run.

2.5 Library construction

The NGS analysis was done by Ms Alvera Vorster at the DNA Sequencer Unit of the Central Analytical Facilities (CAF), Stellenbosch University (http://www.sun.ac.za/english/faculty/science/CAF/units/dna-sequencer). The candidate observed and assisted Ms. Vorster with all the steps of this analysis.

Library construction is the process of producing DNA fragments of a specific size range for NGS sequencing. Firstly, the DNA from each sample is prepared by fragmenting it into smaller pieces so that it is compatible with the sequencing platform. The fragmented DNA goes through multiple steps of purification, amplification, size selection, hybridization and capture by probes designed to be complementary to the target regions. The captured libraries of each individual are pooled into one library that is sequenced on a sequencing platform in a single run. Each individual's library is uniquely barcoded prior to pooling to ensure that the sequencing data is correctly analysed. The SureSelect Target Enrichment System for sequencing on Ion Proton protocol was followed for the construction of our sample's libraries (Agilent Technologies, Santa Clara, MA, USA). Figure 2.1 shows a summary of the library construction process that was followed.

2.5.1 DNA quantification

High-quality genomic DNA is necessary for NGS platforms to produce good standard sequencing data. Therefore, the quality of the DNA in our study was assessed prior to sequencing on an Ion Torrent Proton™ NGS platform (ThermoFisher Scientific, Waltham, MA, USA). The genomic DNA from each sample was quantified using Qubit 1x dsDNA high sensitivity assay kit (ThermoFisher Scientific) and assessed on a Qubit 4.0 fluorometer according to the manufacturer's protocol, MAN0017455 REVB.0. Furthermore, spectrophotometry of the DNA was performed to determine the purity of the samples using the NanoDrop®

ND-1000 (ThermoFisher Scientific). Additionally, to assess if the DNA is intact, the genomic quality scores (GQS) of the samples were determined on the LabChip GXII Touch using the DNA Extended Range LabChip and Genomic DNA Reagent Kit (PerkinElmer, Waltham, MA, USA).

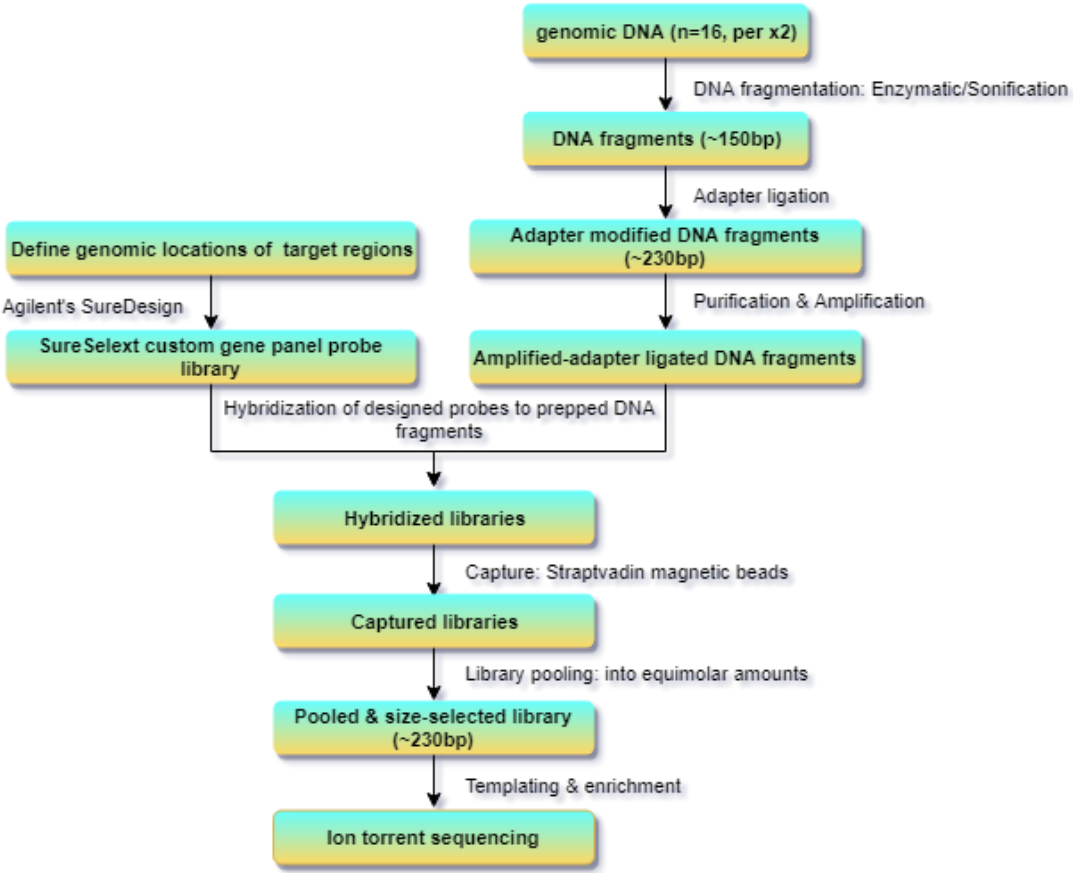


Figure 2.1 Summary of the library construction workflow followed in the present study. The workflow is adapted from the SureSelect Target Enrichment System protocol for ion torrent sequencing (Agilent Technologies, Santa Clara, MA, USA).

2.5.2 DNA shearing or fragmentation

Fragmentation prior to sequencing is necessary to produce DNA fragments of about 150-200 bp, that are compatible with the sequencing platform. Thus, the high-quality DNA from our study participants was sheared into smaller fragments using the enzymatic and physical shearing method. Initially, the enzymatic approach was utilized. The physical shearing was only adopted for DNA samples that failed to produce fragments of optimal size (~150 bp) with the Ion shear enzymes, containing a mixture of shearing enzymes (not specified by the manufacturer).

a. Enzymatic fragmentation

Nineteen of the DNA samples (sixteen from the first sequencing run plus three from the second run) were fragmented with the Ion Shear Plus Reagents (ThermoFisher Scientific) at 37°C for 55 min. Thereafter, the

fragmented DNA was purified and size-selected with the Agencourt™ AMPure™ XP reagent (Beckman Coulter, Brea, CA, USA) as recommended by the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) protocol; Version B0, June 2015. Furthermore, the quality and quantity of the fragmented DNA was assessed on the LabChip® GXII Touch (PerkinElmer, Waltham, MA, USA), using the X-mark chip and HT DNA NGS 3K reagent kit according to the protocol; CLS145098 Rev. E. At this point, only DNA samples with sufficient yield and fragment size distribution were selected for library construction.

b. Physical fragmentation

Physical fragmentation of 13 DNA samples (all were from the second sequencing run) was performed using the Covaris S2 focused ultra-sonicator (Covaris, Inc.; Woburn, MA, USA) using a 20% duty cycle, with 5% intensity, 200 bursts in ten cycles of 60 sec treatment time. Similar to the enzymatically fragmented samples, the DNA fragments were purified and size-selected with the Agencourt™ AMPure™ XP reagent (Beckman Coulter) according to the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) protocol; Version B0, June 2015. However, minor modifications were made to the protocol. These included adding a 1x volume of the Agencourt™ AMPure™ XP reagent to the fragmented DNA and incubating the reaction at room temperature for 5 min. Additionally, the supernatant was collected and transferred to a new 1.5 mL reaction tube, once the beads from the Agencourt™ AMPure™ XP reagent separated from the solution. The bead to sample ratio ensures that DNA fragments of inadequate fragment sizes are discarded during the purification steps. The beads are separated from the solution by placing the tubes containing the mixture on a magnet. Furthermore, a 0.7x volume of the Agencourt™ AMPure™ XP reagent was added to the supernatant and incubated at room temperature for 5 min. Once the beads separated from the solution, 70% ethanol was used twice to wash away any residual impurities. The DNA that was bound to the beads was resuspended in 160 µl low TE (Tris-EDTA) buffer. The resuspended solution was placed on the magnet to separate the beads from the solution and discard them. At this point, the DNA fragments are in the solution and not bound to the beads. The quality and quantity of the fragmented DNA was assessed on the LabChip® GXII Touch (PerkinElmer, Waltham, MA, USA), using the X-mark chip and HT DNA NGS 3K reagent kit according to the protocol; CLS145098 Rev. E. Only DNA samples with sufficient yield and fragment size distribution were selected for library construction.

2.5.3 DNA end-repair of physically fragmented genomic DNA

The ends of the DNA fragments that underwent sonification with the Covaris, were repaired to enable blunt-end ligation. This was performed by adding the end-repair buffer and enzymes from the Ion Plus fragment library kit (ThermoFisher Scientific) in accordance with the manufacturer's protocol; MAN0009847, Revision H.0. An amount of 158 µl of the purified, fragmented DNA sample was incubated for 20 min with 40 µl of the 5x end repair buffer and 2 µl of the end repair enzyme. This was followed by purification with a 1.8x volume of the Agencourt™ AMPure™ XP reagent.

2.5.4 Library preparation

Barcoding of DNA fragments prior to sample pooling is necessary to allow the correct identification of sequencing reads from each sample. Thus, blunt-end ligation of the purified DNA fragments to the Ion Xpress barcode adapters (ThermoFisher Scientific) was performed according to the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) protocol; Version B0, June 2015. Briefly, 25 µl of each sample purified DNA fragments were added to the ligation master mix. The ligation master mix contained 10 µl of the Ion Xpress barcodes 1 to 16.

Following ligation, the barcoded DNA fragments were purified using the Agencourt™ AMPure™ XP reagent and further resuspended in 50 µl nuclease-free water. Additionally, the purified barcoded DNA fragments were amplified for eight cycles using the Herculase II Fusion DNA Polymerase (Agilent Technologies) and the SureSelect Target Enrichment Kit PTN Hyb Module Box #2 according to the manufacturer's instructions; Version B0, June 2015. Amplification of the DNA fragments is necessary to allow the hybridization of the designed probes. Furthermore, the amplified libraries (consists of multiple copies of the adapter-ligated DNA fragments for each sample) were purified with a 1.8x volume of the Agencourt™ AMPure™ XP reagent. Thereafter, 1 µl of each of the purified, amplified libraries were diluted in 9 µl nuclease-free water. The yield and the fragment size distribution of the purified, amplified libraries were analysed with the High sensitivity DNA kit on the Agilent 2100 Bioanalyzer (Agilent Technologies), according to the manufacturer's protocol; G2938-90322 Rev. D.

2.5.5 Library hybridization

Two main library preparation approaches are utilized to enrich the regions of interest, the hybrid capture, and the amplicon-based method (Jennings et al., 2017). The SureSelect target enrichment system utilizes the hybrid-based capture approach. The probes that hybridize and allow the capture of the target regions (the 23 genes associated with monogenic forms of PD) are solution-based, biotinylated RNA oligonucleotide sequences (also referred to as baits) (Kozarewa et al., 2015). The hybridization reaction was performed following the Target Enrichment System protocol with minor modifications (Agilent Technologies, Santa Clara, CA, USA); Version B0, June 2015. The modifications included an incubation step of the reaction for 16 hours at 65˚C and adding 50 µl of MyOne Streptavidin T1 Dynabeads (ThermoFisher Scientific) to the hybridized libraries. Streptavidin contains biotin-binding sites, this enables the capture of the hybridized libraries (John and Quinn, 2008). Subsequently, the hybridized-captured libraries were then purified with the SureSelect Wash 1 and Wash 2 solutions as recommended by the manufacturer, and thereafter, resuspended in 30 µl nuclease-free water. Furthermore, the hybridized-captured libraries were amplified for 11 cycles using the Herculase II Fusion DNA Polymerase (Agilent Technologies) and SureSelect Target Enrichment Kit PTN Hyb Module Box #2 according to the SureSelect Target Enrichment System (Agilent Technologies, Santa Clara, CA, USA) protocol; Version B0, June 2015. This was done to enrich the libraries by producing multiple

copies of the hybridized-captured libraries. Following amplification, the hybridized-captured libraries were purified with a 1.8x volume of the Agencourt™ AMPure™ XP reagents. Thereafter, the yield and fragment size distribution were assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies) using the High sensitivity DNA kit according to the manufacturer's protocol; G2938-90322 Rev. D.

2.5.6 Library size-selection and quantification

Size selection was performed, to ensure that fragments of about 230 bp were retained for sequencing. This enables the elimination of short fragments (e.g. adapter-dimers or primer dimers) and longer fragments (that are not compatible with the sequencing platform) that may influence the sequencing capacity and disrupt the data analysis. Firstly, the libraries were pooled in equimolar amounts (1000 pM) into one tube as recommended by the protocol. Size selection of the pooled library was performed on the Pippen Prep (Sage Science, Beverly, MA, USA) using 2% dye-free Gel cassettes with marker L to retain fragments with a size of about 230 bp. The size-selected library was purified with a 1.8x volume of Agencourt™ AMPure™ XP reagent. Furthermore, it was resuspended in 20 µl nuclease-free water. Ultimately, the yield and fragment size distribution of the purified size-selected library was assessed on the Agilent 2100 Bioanalyzer (Agilent Technologies) using the High sensitivity DNA kit according to the manufacturer's protocol; G2938-90322 Rev. D.

2.6 Templating and sequencing

Following the library construction, the purified size-selected library was diluted to a concentration of 80 pM for template preparation using the Ion PI™ Hi-Q™ Chef Kit (ThermoFisher Scientific) prior to sequencing. Template preparation was achieved by adding 25 µl of the diluted library and loading it onto the Ion Chef liquid handler (ThermoFisher Scientific) for enrichment using the Ion PI™ Hi-Q™ Chef reagents, solutions, and supplies according to the protocol, MAN0010967 REVB.0. Enriched ion sphere particles were loaded onto an Ion PI™ v3 Chip (ThermoFisher Scientific). Once this was done, massively parallel sequencing was performed on the Ion Torrent™ Proton™ system using sequencing solutions, reagents, and supplies according to the protocol, MAN0010967 REV B.0.

2.7 Data analysis

2.7.1 Sequencing data QC and variant calling

The Ion Torrent Suite software (TSS) version 5.4.0 was used to assess the overall performance of the sequencing runs. Quality metrics such as bead loading, Ion Sphere Particle (ISP) density, flow space calibration, base calls, the total number of reads generated, mean read length and Phred quality scores (Q20) were reported. The quality metrics and the sequencing data for each sample is stored in a VCF (Variant Call Format) file format. Mapping of the produced sequence data against the human reference genome (GRCh37/Hg19) was performed using the Torrent Mapping Alignment Program (TMAP) for Ion Torrent Data.

Furthermore, variant calling was achieved using the Torrent variant caller (TVC) plugin, following the software's standard parameters. The TMAP integrates a novel alignment algorithm from multiple aligners such as the Burrows-Wheeler Aligner-MEM (BWA-MEM), Sequence Search and Alignment by Hashing Algorithm (SSAHA) and Super-maximal Exact Matching (Li and Durbin, 2009; Li and Durbin, 2010; Ning et al., 2001; Li, 2012). The TVC plugin identifies single nucleotide variants, multiple nucleotide variants and Indels (insertions and deletions) for each sample from the sequencing data that mapped correctly on the target regions. The VCF files with the called variants for each sample were generated and annotated. Standard mapping quality of the variants from our positive controls was viewed on Integrative Genomics Viewer (IGV) (Robinson et al., 2011).

2.7.2 Variant annotation

VCF files with the called variants for each sample were annotated and combined with SNP & Variation Suite (SVS) package offered by Golden Helix (https://www.goldenhelix.com/products/SNP_Variation/index.html). The following annotation tracks were added, NCBI RefSeq Genes 105 interim v1, 1000 Genomes Phase 3 variant frequencies, NCBI dbSNP 151, ExAC Browser version 3.0 variant frequencies, dbNSFP functional predictions, and scores version 3.0 and the gnomAD exomes and genomes version 2.1.1 variant frequencies. The NCBI Refseq database contains information for Homo sapiens genomes annotated by the NCBI Eukaryotic Genome Annotation Pipeline. This pipeline annotates genes, transcripts, and proteins on multiple completed genome assemblies including the GRCh37/hg19 (Pruitt et al., 2014). The 1000 Genomes Phase 3 dataset contains population frequencies of alleles from reassembled genomes of 2,504 individual's from multiple populations (Consortium The 1000 Genomes Project, 2015). The NCBI dbSNP 151 is the latest version of the dbSNP database that comprises multiple human single nucleotide variants (including insertions and deletions), the publication history and clinical significance associated with the variants (Sherry et al., 2001). The ExAC (Exome Aggregation Consortium) database contains allele frequencies from sequencing data of 60,706 unrelated persons submitted by various disease-specific and genetic studies of multiple populations (Karczewski et al., 2017). The dbNSFP database covers functional predictions from multiple functional predictions *in-silico* tools that give pathogenicity scores of human single nucleotide variants (Liu et al., 2016). Lastly, the genome aggregation database (gnomAD) contains frequency data of variants identified from 125,748 exome sequences and 15,708 whole-genome sequences of unrelated individuals from multiple populations (Karczewski et al., 2019). Figure 2.2 shows a summary of the bioinformatics analysis steps from quality control to variant annotation.

2.7.3 Variant prioritization

The annotated variants were filtered to only include those that are more likely to be disease-causing. Firstly, variants with a minor allele frequency (MAF) of greater than 1% across all populations were filtered out. This

is the recommended MAF filtering threshold to only include rare variants as they are suggested to play a role in the aetiology of complex Mendelian disorders (Frazer et al., 2009). The MAF frequency of the annotated variants we assessed was from the gnomAD (Karczewski et al., 2019) as part of the annotation data. The annotation data also contained pathogenicity predictions from various bioinformatics *in-silico* tools for missense variants. The *in-silico* tools utilized for our study are shown in Table 2.1. These tools investigate the function and phylogenetic conservation of the annotated variants. The frameshift indels were analysed further with Sequence Manipulation Suite (SMS) (https://bioinformatics.org/sms/), to assess if they introduce a premature stop codon.

Figure 2.2 A summary of the bioinformatics pipeline followed to sequence variants. Initial quality control was conducted using the Ion Torrent Suite Software. Furthermore, the sequencing variants were alignment and variants were called with a developed pipeline.

Additionally, we utilized clinical evidence from the ClinVar database to assess if the variants have been previously implicated in the pathogenesis of PD (Landrum et al., 2018). Furthermore, the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines were adopted to further discuss the prioritized variants (Richards et al., 2015). The guidelines classify variants as pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. The criteria for the ACMG variant classification are attached in Appendix 5. Furthermore, we searched the literature and used the PDMutDB database to assess if the variants are previously known to be pathogenic. The database contains information about all known variants in the following PD genes: *SNCA, LRRK2, PRKN, PINK1, DJ-1, FBXO7*.

Table 2.1 *In-silico* tools utilized to assess the pathogenicity of the missense variants and indels

Tool	Category	Reference
SIFT	Functional prediction	(Kumar et al., 2009)
Polyphen-2	Functional prediction	(Adzhubei et al., 2010)
FATHMM	Functional prediction	(Shihab et al., 2014)
MutationTaster	Multi-data integration	(Schwarz et al., 2010)
CADD	Multi-data combination	(Kircher et al., 2014)
DANN	Multi-data combination	(Quang et al., 2015)
LRT	Multi-data combination	(Chun and Fay, 2009)
MetaSVM	Multi-data combination	(Dong et al., 2015)
MetaLR	Multi-data combination	(Dong et al., 2015)
M-CAP	Multi-data combination	(Jagadeesh et al., 2016)
GERP++	Conservation prediction	(Davydov et al., 2010)
PROVEAN	Functional prediction	(Choi and Chan, 2015)
SiftIndel	Insertions and deletions	(Sim et al., 2012)
PROVEAN	Insertions and deletions	(Choi and Chan, 2015)
SMS	Protein translation	(Alzohairy, 2014)

2.8 Design of Polymerase Chain Reaction (PCR) primers and PCR methodology

Primers previously designed with the *Primer 3* software version 4.0.0 (http://primer3.ut.ee/) were utilized to validate the presence of the top prioritized variants (Koressaar and Remm, 2007). Sequence data for designing the primers were accessible from the Ensembl Genome Browser database (http://www.ensembl.org). In addition, the Primer-Basic Local Alignment Search Tool (BLAST) (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to validate the specificity of the primer binding positions. The designed primers were synthesised by Integrated DNA Technologies (IDT, Coralville, Iowa, USA) and were supplied by Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, SA. The primer sequences, PCR conditions and the expected band size of the PCR products are attached in appendix 6.

For the PCR analysis of the prioritized *PINK1* exon 4 and *PRKN* exon 5 variants, the exons were amplified from the genomic DNA of study participants carrying the variants and family members for whom we had DNA. The total volume of the PCR reaction was 25 µl, containing 4 µM of each primer (forward and reverse), 75 µM of dNTPs (GeneDireX, Taiwan), 1.5 mM $MgCl_2$ (Promega, USA), 1 unit of GoTaq DNA polymerase, 1X Green GoTaq Flexi Buffer (Invitrogen, USA) and 200 ng of genomic DNA. A non-template control was included in each PCR run to check for any contaminations in the reaction mix. The PCR amplification occurred in an ABI 2720 Thermocycler (Applied Biosystems Inc., Foster City, California, USA). The PCR conditions for each variant are shown in appendix 6. Briefly, an initial denaturation step was set up at 94˚C for 5 min. This was followed by cycling for 35 cycles at 55˚C to allow the primers to anneal and a 3 min extension step. Lastly, the reaction was extended at 72˚C for 7 min. Once the reaction was done, the tubes were left in the machine on hold to cool at 4˚C until they were used for further analysis.

2.9. Agarose gel electrophoresis

Once the PCR amplification was performed the DNA fragments were assessed by agarose gel electrophoresis. This is done to ensure that the correct region of interest is amplified and that there was no contamination in the PCR reaction mix. 1% agarose gel was used which involved dissolving 1 g of agarose in 100 mL 1x Tris-acetate-EDTA (TAE) buffer. The solution was allowed to dissolve in a microwave for 2 min and left to cool down at room temperature. Once the solution reached approximately 50˚C, 4 µl of Ethidium bromide (EtBr) (Sigma-Aldrich., United States) stain was added. EtBr binds to the DNA allowing the visualization of the PCR products under ultraviolet light. Approximately 5-8 µl of each PCR reaction was loaded on the prepared gel. Additionally, a molecular marker that can be used to track DNA fragments that range from 100-3000 bp was also loaded into the prepared gel (Invitrogen, USA). The electrophoresis was run at a velocity of 120 V for 30 min. The SynGene UV gel system (Synoptics Ltd., Cambridge, UK) and the GeneTools software version 3.0.6 (Synoptics Ltd., Cambridge, UK) was used to visualize the amplified PCR DNA fragments.

2.10 Post PCR purification

The produced PCR DNA fragments were further purified to remove excess dNTP's, primers, enzymes and components of the buffer that are not required for Sanger sequencing. Briefly, 8 µl of the PCR reaction was mixed with one unit of Shrimp alkaline phosphatase (SAP) (Cleveland, Ohio, USA) and one unit of Exonuclease (Exo I) (Promega, USA). The reaction was placed in an ABI 2720 Thermal cycler under the following conditions, at 37˚C for 15 min, at 80˚C for 15 min to deactivate the enzymes (Applied Biosystems Inc., Foster City, California, USA).

2.11 Variant validation using Sanger sequencing

Sanger sequencing of the amplified DNA fragments was performed at the Stellenbosch University CAF (Stellenbosch, South Africa). DNA sequencing at the facility was conducted using the BigDye Terminator V3.1 sequencing kit (Applied Biosystems, CA, USA) following the optimized in-house method. The sequence analyses software, BioEdit version 7.0.5.3 was utilized to analyse the sequences (Hall, 1999). Reference sequences of the amplified exons were used to align the sequencing data and were accessed from the Ensembl Genome Browser database (https://www.ensembl.org/index.html).

2.12 Mutation screening of *GBA*

GBA is an important risk factor for PD with a number of pathogenic mutations that have been identified (https://www.ncbi.nlm.nih.gov/books/NBK1269/). We used a nested PCR method to screen this gene. Previously designed *GBA* specific primers were used to amplify the gene (Stone et al., 2000). The primers were synthesised by Integrated DNA Technologies (IDT, Coralville, Iowa, USA) and were supplied by Inqaba Biotechnical Industries (Pty) Ltd, Pretoria, SA. The first PCR primers amplify the gene into three fragments (fragment 1 [exon 1-5], fragment 2 [exon 5-7] and fragment 3 [exon 8-11]). The reaction contained a total volume of 25 µl of which 4 µM of each primer (forward and reverse), 75 µM of dNTPs (GeneDireX, Taiwan), 1.5 mM MgCl2 (Promega, USA), 1 unit of GoTaq DNA polymerase, 1x Green GoTaq Flexi Buffer (Invitrogen, USA) and 200 ng of genomic DNA was added. This was followed by gel electrophoresis and PCR purification as described in sections 2.9 and 2.10. Subsequently, 1 µl was taken from this reaction and used as a DNA template for the second PCR amplification with primers that amplify the fragments into individual exons (exon 1 to 10). The same PCR amplification protocol, gel electrophoresis, and PCR purification steps were followed. Additionally, the frequency of novel *GBA* variants were assessed in 110 ethnic matched controls. The same nested PCR approach was also utilized for screening of all control individuals. The PCR conditions and primer information are shown in appendix 6.

CHAPTER 3: RESULTS

Part A: Gene Panel

3.1 Selection of study participants for sequencing runs

We selected 24 PD patients for mutation screening using the Agilent SureSelect Target Enrichment custom-designed gene panel. The patients were analysed on 2 separate runs. Run 1 comprised of 12 patients and 4 positive controls; whereas run 2 comprised of 12 patients, 1 positive control and 3 negative controls (Appendix 3). The positive controls were PD patients in whom disease-causing mutations had previously been found by members of our research group, using established methods such as Sanger sequencing and MLPA. These mutations included two non-synonymous SNV's (*PRKN* p.G430D and *LRRK2* p.G2019S), three CNV's (*SNCA* triplication, *PRKN* exon three and four deletions), and a *PRKN* frameshift 40 bp deletion in exon three (p.P113Tfs*). These were included to assess the sensitivity of the custom gene panel to detect these known mutations.

Once we had established that the known SNV's and the 40 bp deletion were successfully detected, the second sequencing run was performed. During our first sequencing run, we were unable to detect the CNV's in the positive controls. Thus, the individual with the *SNCA* triplication and the three negative controls (WT) were included in the second run to detect the difference between a normal gene copy and a gene dosage mutation.

3.2 PD gene panel design

The candidate PD genes included in our custom-designed gene panel were selected by performing an extensive literature search and 23 genes, that met our criteria, were selected. Our selected gene targets were associated with autosomal dominant PD (*SNCA, LRRK2, VPS35, GCH1, DNAJC13, TMEM23, HTRA2, GIGYF2, CHCHD2, EIF4G1, RIC3, and GBA),* autosomal recessive PD (*PRKN, PINK1, DJ-1, ATP13A2, PLA2G6, FBXO7, DNAJC6, SYNJ1, VPS13C, SLC6A3*) and one gene known to cause the X-linked inherited form of PD *(RAB39B)* (Appendix 7).

The custom panel was designed using SureDesign Agilent's web-based application for designing target enrichment libraries. SureDesign allowed us to define our target genes by accessing the Ensembl's database (https://www.ensembl.org/index.html) for the gene transcripts. The gene panel library consisted of a total of 1,518 probes covering 484 exons and represented 112,387 bp of the coding region of the target genes. The probes were designed to be complementary to the exonic regions of our target genes, extending 10 bp into the intronic regions on both sides of the exons. The probes were successfully designed with the Agilent SureDesign default settings for almost all of the gene targets except for *GIGYF2, PINK1, PLA2G6,* and *TMEM230*. These genes contained regions that were difficult to enrich (multiple repeats and high GC content regions). To overcome this problem, the masking of repetitive sequences for *GIGYF2, PINK1*, and *PLA2G6* was

set to be less stringent while the probe density for *TMEM230* was doubled. This resulted in all of the target regions being potentially covered by the designed probes. A representation of the improvement of coverage after changing the parameters is shown in Appendix 7.

3.3 Library construction quality control (QC)

3.3.1 Sample QC

The SureSelect Target enrichment system requires that the quality of the DNA is assessed prior to sequencing on the Ion Proton platform. Firstly, the Qubit 4.0 Fluorometer dsDNA high sensitivity kit was used to quantify the amount of genomic DNA in each sample. Furthermore, the quantity and integrity of the DNA were assessed by spectrophotometry using the Nanodrop and electrophoresis on the PerkinElmer LabChip. High-quality dsDNA has an OD 260/280 ratio of between 1.8 and 2.0. An OD 260/280 ratio ranging from 1.82-1.99 was obtained for all our samples. For successful library construction, ≥ 1000 ng of genomic DNA is required. The determined amount of the genomic DNA of all our samples was found to be > 1000 ng. Intact high molecular weight genomic DNA is also essential, and it is determined by a genome quality score (GQS) between 3 and 5. The GQS score of our DNA samples ranged from 3-4.5.

3.3.2 DNA fragmentation QC

The DNA samples were fragmented during the library construction steps. This is necessary to produce optimal size DNA fragments for the sequencing platform. Firstly, the enzymatic fragmentation method was performed using the Ion Shear Plus Reagents containing a mixture of enzymes not specified, able to cut dsDNA into fragments with different types of ends (recessed, overhang or blunt ends). Following fragmentation and purification, the quantity and quality of the produced DNA fragments were assessed. The concentration of the DNA fragments was determined by integrating the area under the base peak curve of an electropherogram produced on the PerkinElmer Labchip. As recommended by the protocol, the average fragment size should be approximately 130 bp and it must show a distribution between 50 and 250 bp. Only 19 of our samples showed sufficient fragment size distribution and the average fragment size ranged from 126-147 bp. Figure 3.1 shows the Labchip electropherogram of three samples sheared with the Ion shear enzymes. To overcome this problem, samples that failed to produce sufficient size DNA fragments were sheared using the sonification method on the Covaris. The fragment size distribution of the samples sheared with the Covaris was adequate (50-250 bp), with an average fragment size between 130 and 150 bp. Figure 3.2 shows the Labchip electropherogram of three samples sheared with the sonication method. The improvement in fragmentation of one DNA sample that failed with the enzymatic shearing versus the Covaris is shown in Figure 3.3.

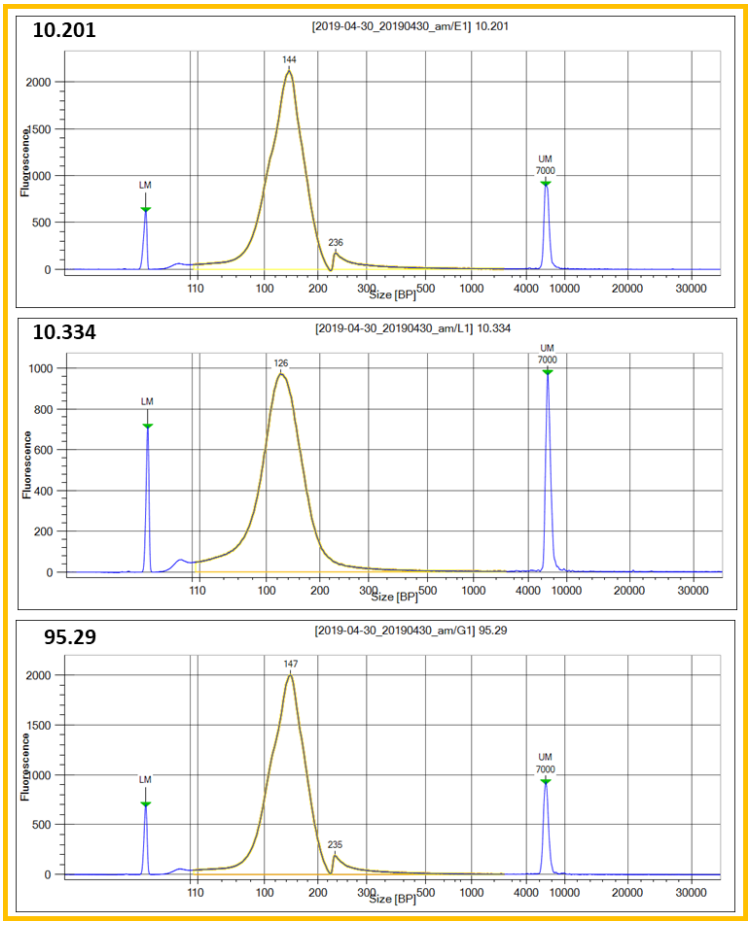


Figure 3.1 Labchip electropherograms showing the size distribution of DNA fragments sheared with the Ion shear enzymes for three representative DNA samples. The X-axis represents the base pairs while the Y-axis shows the intensity (fluorescence) of the fragments. The upper and the lower size markers are represented by the labels LM and UM, respectively. The average fragment size for sample 10.201 was 144 bp, for sample 10.334 it was at 126 bp and it was at 147 bp for sample 95.29

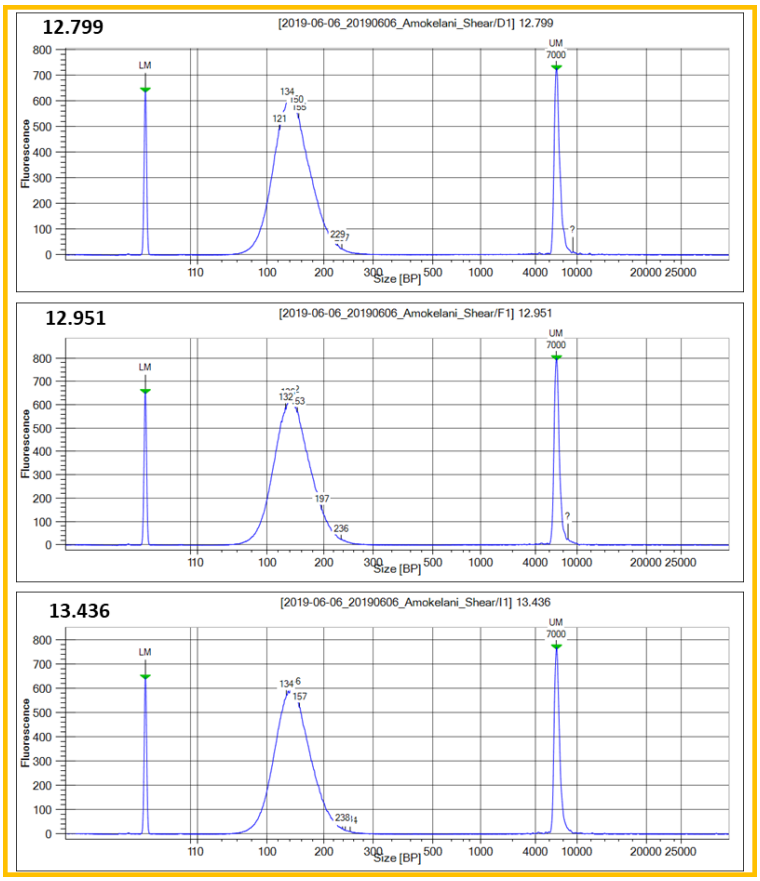


Figure 3.2 Labchip electropherograms showing the size distribution of DNA fragments of three representative DNA samples sheared with the sonication method on the Covaris. The X-axis represents the base pairs while the Y-axis shows the intensity (fluorescence) of the fragments. The upper and the lower marker are represented by LM and UM, respectively. The average fragment size for sample 12.799 was 134 bp, for sample 12.951 it was at 132 bp and 134 bp for sample 13.436.

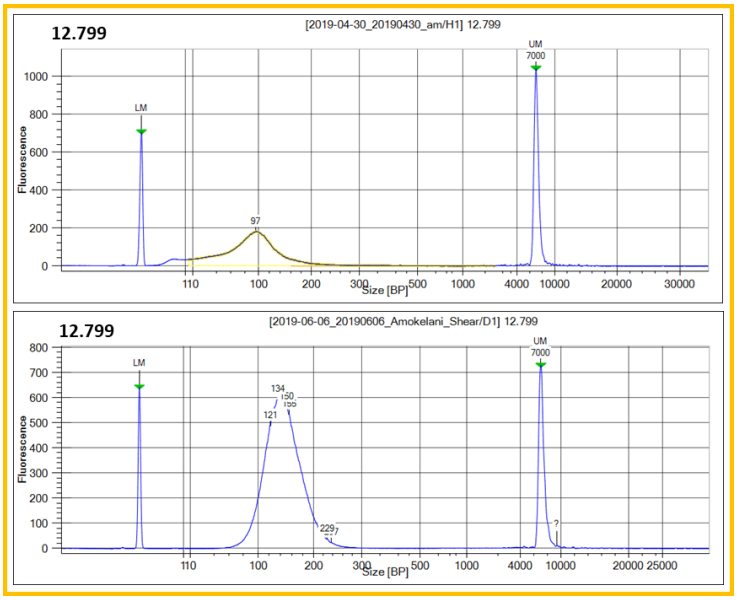


Figure 3.3 Labchip electropherograms showing the difference in the size distribution of DNA fragments of sample 12.799, sheared with the enzymes versus the Covaris. The X-axis represents the base pairs while the Y-axis shows the intensity (fluorescence) of the fragments. The upper and the lower marker are represented by LM and UM, respectively. The top panel was generated with the enzymes (which was unsuccessful) and the bottom panel is with the sonication method (which was successful).

3.3.3 Ligated library QC

The DNA fragments of each sample were ligated to the barcode adapters. Each sample's sheared, barcoded, ligated and amplified DNA fragments are referred to as a library. The quality and quantity of the amplified barcoded libraries were assessed to ensure that the adapters are ligated to the ends of the DNA fragments. The electropherogram of the adapter-ligated libraries showed an average fragment size of 227.5 bp (range 216-234 bp). The concentration of the ligated libraries was also determined by integrating the area under the base peak curve. Figure 3.4 shows the Agilent 2100 Bioanalyzer electropherogram size distribution of the adapter-ligated library of samples 10.201 and 10.334.

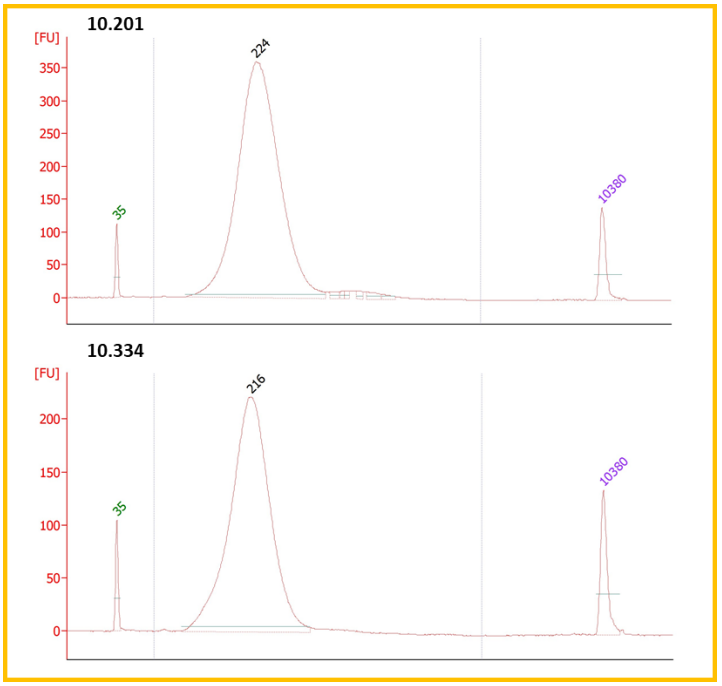


Figure 3.4 Bioanalyzer electropherogram showing the fragment size distribution of the ligated libraries of two representative samples 10.201 and 10.334. The X-axis represents the base pairs while the Y-axis shows the intensity (fluorescence) of the fragments. The lower and the upper size markers are represented by the labels 35 bp and 10380 bp, respectively. The average fragment size for both samples increased after ligation, indicating that the adapters were successfully added.

3.3.4 Hybridized, captured and size-selected library QC

The amplified, adapter-ligated libraries were then hybridized to the designed probes. The libraries that successfully hybridized were captured by streptavidin beads and amplified to enrich the captured regions. The hybridization reaction requires 750 ng of prepared DNA in a 3.4 µl volume (initial concentration of 221 ng/µl) as recommended by the protocol. For our study, the concentration of our prepped library ranged between 279 and 750 ng. A vacuum concentrator was used to concentrate samples with low amounts of library. The quantity and quality of the hybridized-captured libraries were assessed with the Agilent 2100 Bioanalyzer. The protocol recommends that the libraries should be pooled in equimolar amounts in such a way that each barcoded sample is present in equal quantities. The pooled library is then size-selected by electrophoresis on a 2% agarose gel (E-Gel system). The hybridized-captured libraries were pooled in equimolar amounts (1000 pM) and size-selected to retain DNA fragments of ~230 bp. The quantity and quality of the size-selected library were also analysed on an Agilent 2100 Bioanalyzer. The fragment size distribution of the pooled library showed an average fragment size of ~239 bp (Figure 3.5).

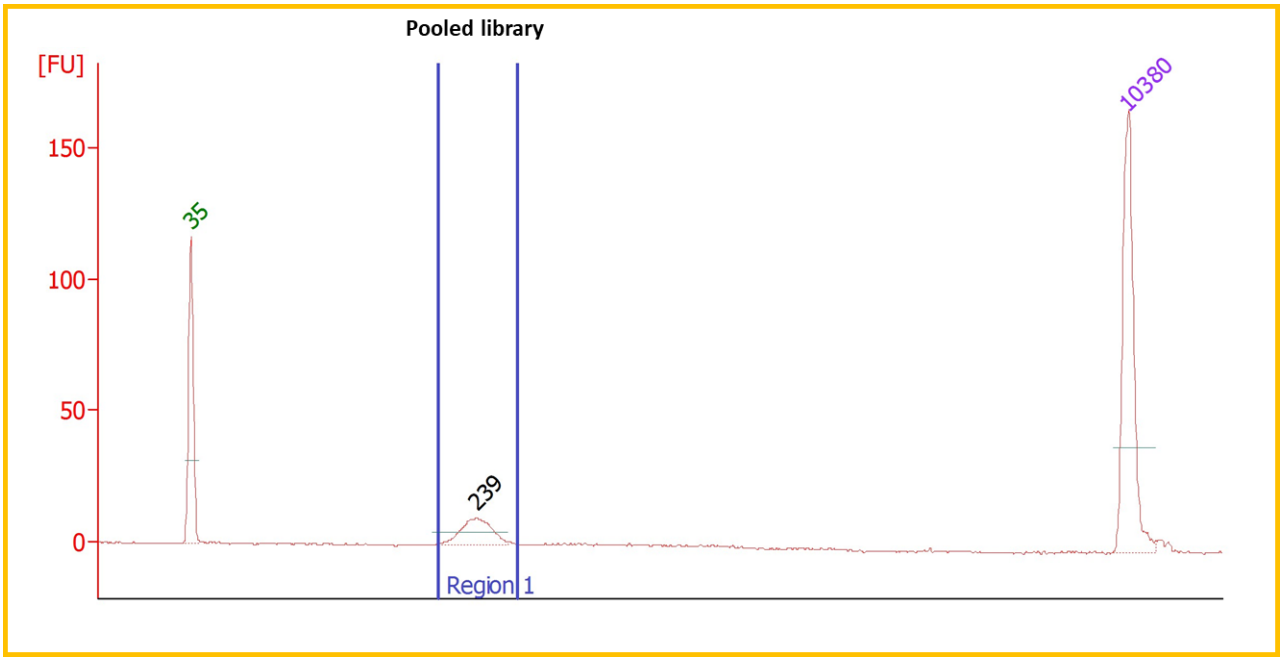


Figure 3.5 Bioanalyzer electropherogram showing the fragment size distribution of the pooled (all 16 samples) and size-selected library for the first sequencing run. The X-axis represents the base pairs while the Y-axis shows the intensity (fluorescence) of the fragments. The lower and the upper size markers are represented by the labels 35 bp and 10380 bp, respectively. The average fragment size of the pooled and size-selected library for both sequencing runs was optimal (~230 bp).

3.4 Ion Torrent sequencing, alignment, and coverage analysis

Template preparation and enrichment of the size-selected library was performed using the Ion Chef. This was followed by sequencing on the Ion Proton platform at CAF. For the first sequencing run, a total of 84,022,874 raw reads of which 64% were usable, were generated. The run resulted in a total of 11G bp of sequencing data. The generated reads had a mean read length of 138 bp (range 136-141). Only high-quality sequenced bases (with a Phred quality score of Q20 or greater) in a form of a VCF file format were included for alignment to the reference genome and variant calling. The Phred quality score of Q20 indicates that 99% of the bases that are included in the analysis were identified correctly by the sequencer. The average coverage on target regions per sample was 554.7x (range 344.7-802.9). The sequencing reads were aligned to the GRCh37/hg19 reference genome. Only reads that mapped to our target regions were investigated. Sequencing reads that were mapping off target were not analysed to minimize complications with the data analysis. Although most of the reads were found to map off-target (greater than 70%), we had adequate coverage (>30x) on our target regions for variant calling. The sequencing run yielded a mean base call uniformity of greater than 99.3% across all samples. The sequencing and coverage quality metrics are shown in Table 3.1.

The same library preparation protocol and sequencing platform utilized for the first sequencing run were also adopted for the second run. The second run yielded a total of 100,767,358 raw sequencing reads of which 71% were usable. The run generated a total of 16.5G bp of sequencing data, and the reads had a mean length of 141 bp (range 138-143). Sequenced bases with a Phred quality score of Q20 or greater were filtered for further analysis. The average coverage on target regions per sample was 429.4x (range 269.1-609). Like the first sequencing run, the majority of our sequencing reads mapped off-target (about 70%). However, we had sufficient coverage (>30x) for variants called on our target regions. The sequencing run yielded a mean base call uniformity of greater than 99.2% across all samples. The sequencing and coverage quality metrics for run 2 are shown in Table 3.2.

The number of reads generated, on-target percentage and average read depth of sample 10.344 (a positive control that was included in both runs) were assessed. More reads were produced (5,208,008 reads vs 2,492,152 reads) and the on-target percentage (23.13% vs 17.13%), as well as the mean read depth (522.2x vs 269.1x) of the sample, was better for the first run. Thus, indicating that our first sequencing run performed better than the second one (Table 3.1 and Table 3.2).

Table 3.1 Quality metrics of each sample for the first run

Sample ID	[1]Number of reads generated	On target percentage (%)	Average read depth[2]	Mean read length (bp)	[3]Number of variants
88.28	4,215,944	26.55	561.9	138	84
94.69	5,428,526	27.87	802.9	141	102
74.53	4,626,515	29.51	607.4	139	87
90.95	5,282,946	23.85	583	140	79
10.334*	5,208,008	23.18	522.2	138	73
10.322	5,781,692	23.23	585.1	138	98
89.01	4,651,278	24.09	524.7	139	72
66.18	5,721,364	25.55	592	139	70
59.91	4,366,349	27.19	545.9	139	67
12.819	5,318,348	17.49	413.3	136	94
10.783	6,470,038	26.46	637.4	138	70
12.726	5,354,357	16.07	344.7	136	76
78.74*	6,362,646	22.01	563.3	138	114
11.861	4,135,309	25.50	471.8	137	73
82.16*	4,267,527	26.72	477.8	138	69
11.927*	5,159,138	25.55	641.9	139	86

[1]Reads with a Phred quality score ≥ Q20; [2]Average read depth per sample (indicates how many reads, on average are aligned at a given reference base position); [3]variants called in each sample; * positive control

Table 3.2 Quality metrics of each sample for the second run

Sample ID	Number of reads generated	On target percentage (%)	Average read depth	Mean read length (bp)	Number of variants
13.378	6,926,673	15.22	609	142	76
12.731	4,009,668	16.44	451	142	79
13.272	3,511,300	19.97	479	143	97
12.799	3,973,567	19.89	547.9	142	83
10.334*	2,492,152	16.55	269.1	138	72
13,164	4,040,911	16.97	461	142	66
12.951	3,835,365	16.42	439.3	141	90
95.63	4,082,226	14.25	368	141	88
13.435‡	3,687,733	17.74	432	141	83
13.436‡	3,364,547	16.35	404.4	141	84
13.437‡	3,825,188	11.92	343.4	141	78
78.84	3,959,927	13.75	353.2	141	67
81.58	3,516,602	13.92	306.6	142	85
84.30	3,450,649	14.48	319.1	141	77
10.201	3,725,257	20.41	555.5	141	66
95.29	3,636,152	20.74	531.3	141	78

[1]Reads with a Phred quality score ≥ Q20; [2]Average read depth per sample (indicates how many reads, on average are aligned at a given reference base position); [3]variants called in each sample;* positive control; ‡negative control

3.5 Identification of sequence variants

As shown in Tables 3.1 and 3.2, an average of 82 variants (range 67-114) per sample was identified for our first sequencing run and for the second run, an average of 79 variants (range 66-97) was detected. The variants called from the data of both our sequencing runs were annotated with SNP & Variation Suite (SVS) package offered by Golden Helix. Annotation tracks that give additional information about the variant location, gene function, gene ontology, disease association, population frequency, pathogenicity, and clinical significance were added. Although the capture probes were designed to target the exonic regions of the candidate genes, they also extended to about 10 bp into the intronic regions to sequence the splicing sites.

Overall, from the combined annotation data of the 12 patients from the first run and the 12 patients from the second run, most of the variants were found to be exonic (186) followed by intronic (15) variants (Figure 3.6). Additionally, one 5' UTR and one 3' UTR variant were detected. For our study, we focused on exonic variants including those that may alter splicing as they are more likely to interfere with the protein function and cause disease.

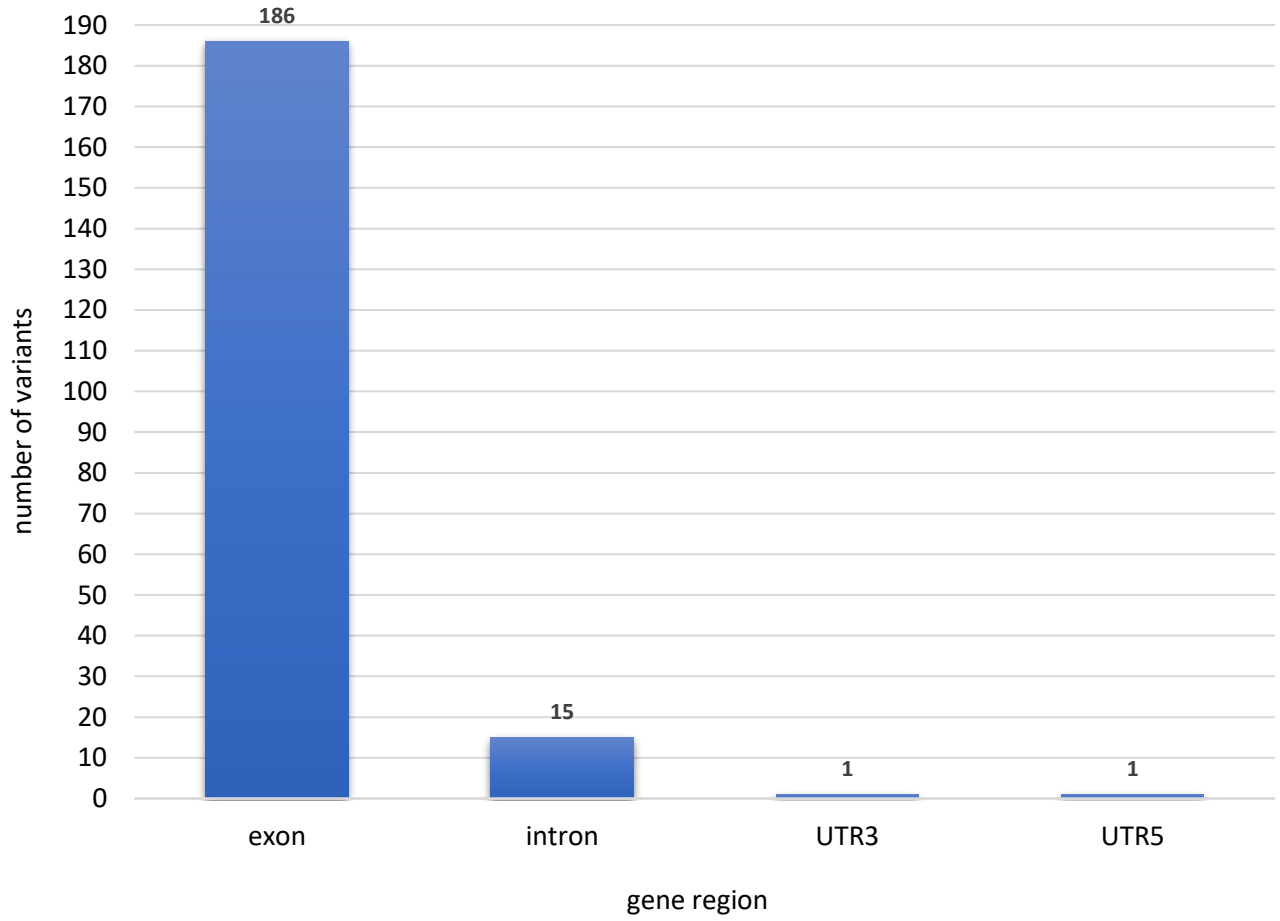


Figure 3.6 Types of variants identified. The majority of the variants identified from the annotation data of the 24 patients were exonic (n=186). As expected very few intronic (n=15), 5'UTR (n=1) and 3'UTR (n=1) variants were detected as our gene panel was designed to mostly screen for exonic regions and the exon and intron boundaries.

3.5.1 Exonic variants

The exonic variants were annotated as either missense, synonymous, frameshift, inframe deletion and insertion, initiator codon, and stop gained. The majority of the exonic variants were missense (90), followed by synonymous (81) (Figure 3.7).

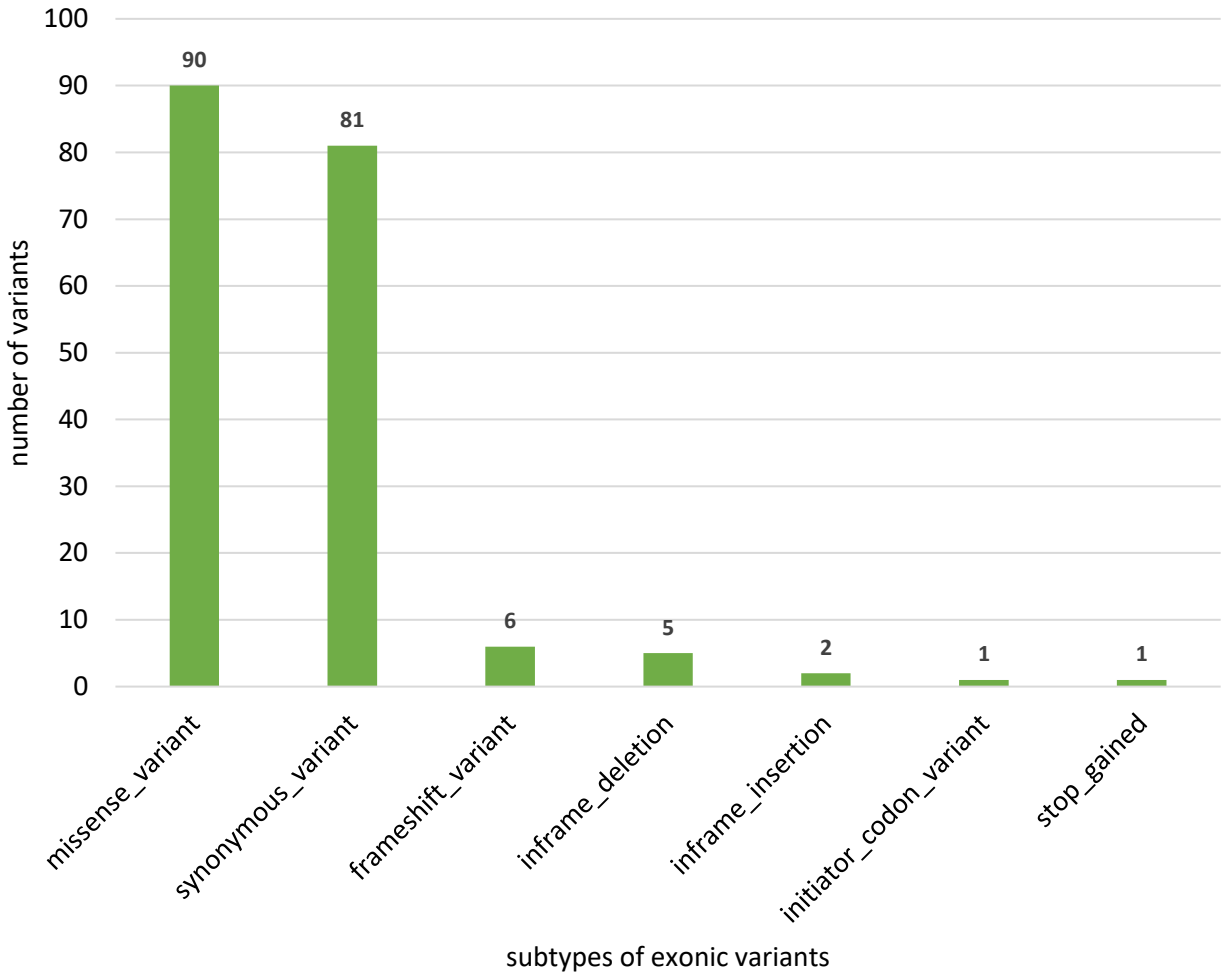


Figure 3.7 Sub-types of exonic variants identified. The majority of the detected exonic variants from our annotation data of 24 patients were missense (n=90) followed by synonymous (n=80). Few of the exonic variants were found to be frameshift variants (n=6), inframe deletions (n=5) and insertions(n=2). Only one initiator codon and a stop gained variant were identified.

Additionally, we assessed which of our candidate genes contained the most exonic variants. Thus, we found that most occurred in the *VPS13C* (31) gene followed by *LRRK2* (22) and *ATP13A2* (19) (Figure 3.8). We did not identify any exonic variants in the *CHCHD2*, *GCH1* and *SNCA* genes. The *VPS13C* gene harboured predominantly missense variants while synonymous variants frequently occurred in the *ATP13A2* gene (Figure 3.9). The inframe deletions were only found in the *GIGYF2* gene. Furthermore, the inframe insertions occurred only in *SYNJ1* and *GIGYF2*. Lastly, the initiator codon variant and stop gained variant were each only found in *FBXO7* and *ATP13A2,* respectively.

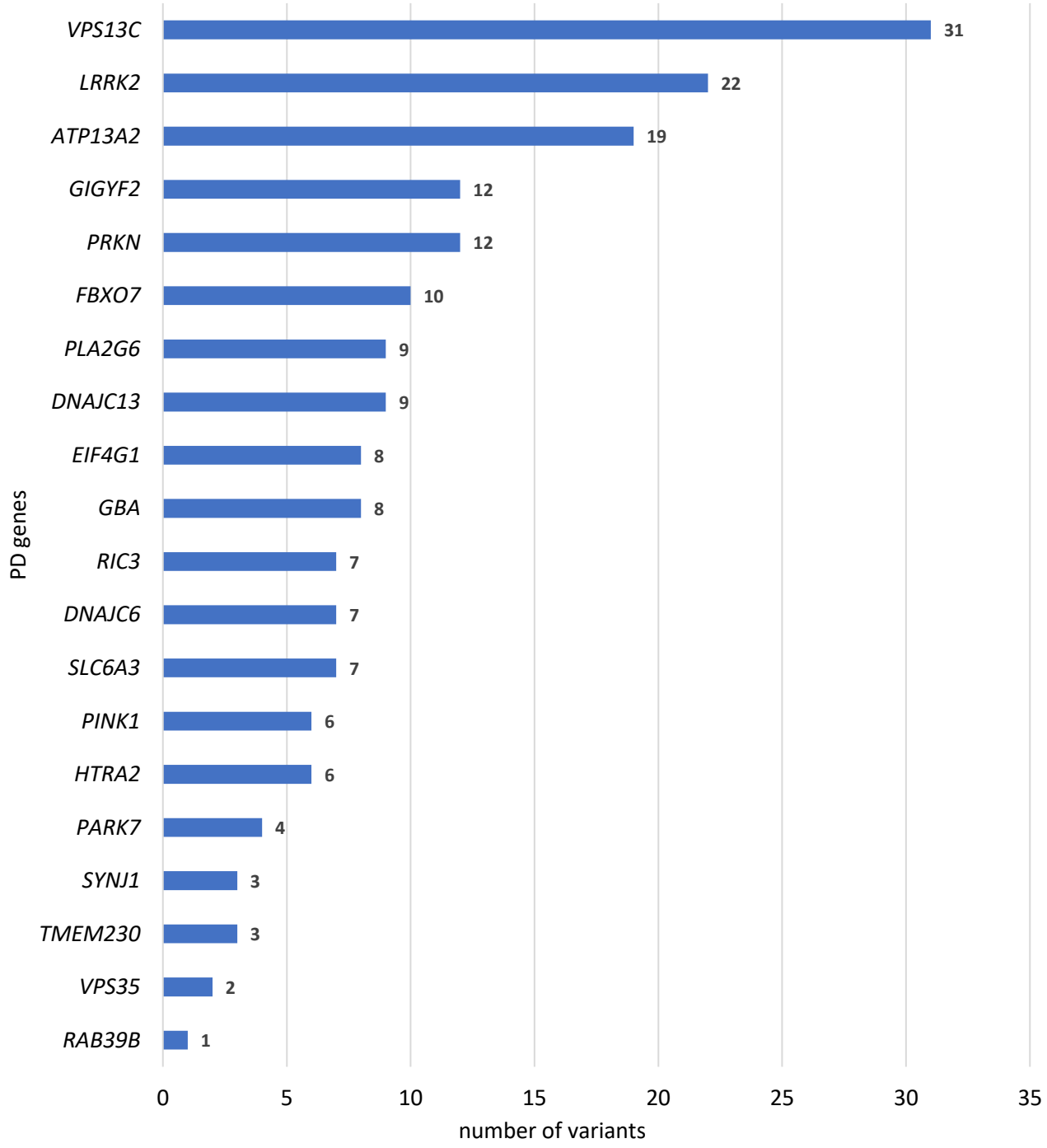


Figure 3.8 Number of exonic variants identified per gene. The majority of the exonic variants from the annotation data of 24 patients were found in the *VPS13C* gene (n=31) followed by *LRRK2* (n=22) and *ATP13A2* (n=19). Few exonic variants were identified in *GIGYF2* (n=12), *PRKN* (n=12), *FBXO7* (n=10), *PLA2G6* (n=9), *DNAJC13* (n=9), *EIF4G1* (n=8), *GBA* (n=8), *RIC3* (n=7), *DNAJC6* (n=7), *SLC6A3* (n=7), *PINK1* (n=6), and *HTRA2* (n=6). The least exonic variants occurred in *PARK7* (n=4), *SYNJ1* (n=3), *TMEM230* (n=3), *VPS35* (n=2) and only one variant was detected in the *RAB39B* (n=1) gene. No exonic variants were found in *SNCA*, *CHCHD2,* and *GCH1*.

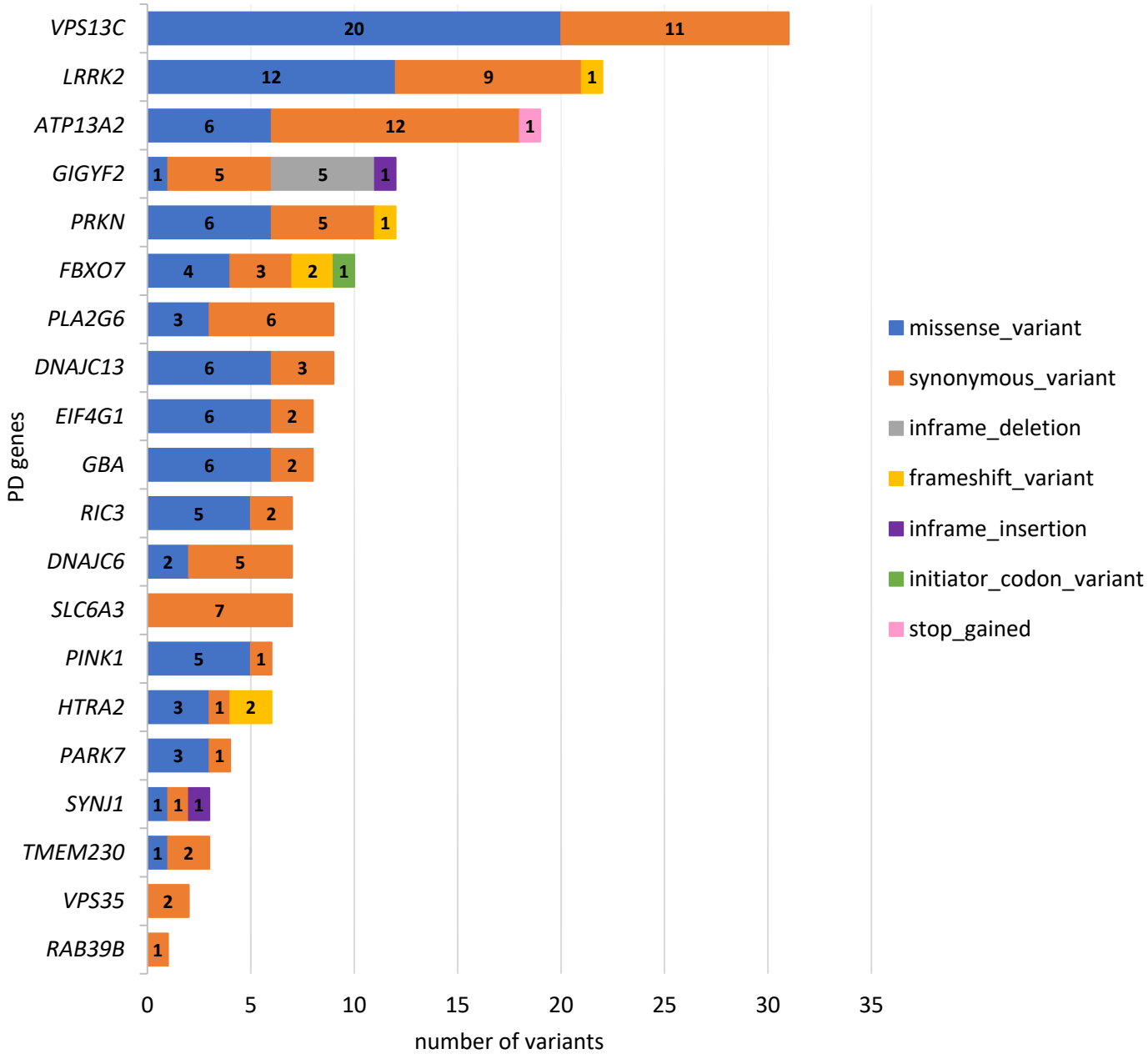


Figure 3.9 Sub-type of exonic variants identified per gene. The majority of the missense variants from the annotation data of 24 patients occurred in the *VPS13C* gene. The synonymous variants mostly occurred in *ATP13A2.* Only synonymous variants were found in *VPS35* and *RAB39B*. The frameshift variants were only detected in *LRRK2*, *PRKN*, *FBXO7,* and *HTRA2*. Furthermore, the one inframe insertion only occurred in *SYNJ1*.

3.5.2 Known exonic variants

a. Positive controls

We successfully detected the presence of the two nonsynonymous mutations (*PRKN* p.G430D and *LRRK2* p.G2019S) and the 40 bp deletion in *PRKN* (p.P113Tfs*). The *PRKN* 40 bp deletion was correctly identified by the variant calling software and it was accurately annotated. Figure 3.10 shows an IGV screenshot of the 40 bp deletion. Therefore, the gene panel was successfully designed as we were able to identify the validated mutations in some of our positive controls. Also, the sequencing reads for these variants were of good quality.

Figure 3.10 A screenshot of the Integrative Genomics Viewer showing the *PRKN* 40 bp deletion resulting in a frameshift mutation (p.P113Tfs*) detected by our gene panel from one of the positive controls. The variant is visualized by importing the BAM files (contains chromosomal locations of the sequencing reads) into IGV.

Additionally, we set out to detect the previously identified CNV's from our other positive controls i.e. the *SNCA* triplication, and the *PRKN* exon 3 and exon 4 deletions. These variants could not be detected in our first run as we had not included WT controls. Therefore, in the second run, we included three WT controls and one positive control to detect a CNV change. However, these variants could not be identified with the annotation pipeline we used as a CNV specific pipeline is required. Unfortunately, as we were unable to find someone who could assist us with this type of analysis, we did not focus on CNV detection in the present study. However, the raw data is available for future analyses for the identification of this type of variant.

b. Previously identified exonic variants

Overall, from the combined sequencing data of the 24 individuals, we identified four variants that have been annotated as pathogenic or potentially pathogenic on the ClinVar database. ClinVar is a widely used database that contains a collection of information about the clinical significance of variants, in accordance with the ACMG guidelines, that were submitted by various laboratories. These were all *GBA* variants, the p.L483P (p.L444P) and the p.R170L (p.R131L) variants are reported as pathogenic, the p.D179H (p.D140H) is reported as a likely pathogenic/variant of uncertain significance and the p.E365K (p.E326K) variant is reported as likely pathogenic/risk factor/benign variant. Additionally, eight other known variants *(ATP13A2* [p.V1133V, p.G158R], *LRRK2* [p.M2397T, p.I610T], *PLA2G6* [p.V699V, p.A324A, p.S258L], *PRKN* [p.E310D]) which were previously classified as variants of uncertain significance (VUS) on ClinVar were also found in our annotation data of the 24 PD patients.

3.5.3 Prioritized rare and novel exonic variants

Rare and novel missense, frameshift and inframe indels were prioritized in our study as they are expected to alter the protein sequence and impact the function of the gene. Firstly, rare variants were selected based on a MAF of less than 0.01 (the gnomAD database was used to assess the variant frequency). Furthermore, novel variants that were not reported in public databases were chosen. Overall, 54 rare and novel variants plus the known *GBA* risk factor were selected (Appendix 8). Subsequently, these variants were further prioritized using the following selection criteria:

1. Rare and novel missense variants predicted to be pathogenic by at least six of the 11 *in-silico* tools:
    a. SIFT
    b. Polyphen-2
    c. FATHMM
    d. MutationTaster
    e. CADD
    f. DANN
    g. MetaSVM
    h. MetaLR
    i. LRT (likelihood ratio test)
    j. PROVEAN
    k. predicted to be conserved by a GERP++
2. Rare and novel frameshifts predicted to be pathogenic by both SIFT-indel and PROVEAN.

This resulted in 14 variants being prioritized and their pathogenicity predictions are shown in Table 3.3. In addition, Table 3.3 includes the clinically significant *GBA* variant p.E365K, that did not meet the above-mentioned inclusion criteria. Therefore, in total 15 variants were prioritized.

Table 3.3 Prioritized rare, novel missense frameshift, and the known *GBA* risk factor

PD genes	*DNAJC13*	*FBXO7*		*PINK1*		*PRKN*	*VPS13C*			*RIC3*	*GBA*				
Inheritance	AR	AR		AR		AR	AR			AD	AD				
Variant	c.5630G>A p.R1877Q	c.948_949delTC p.L317Gfs	c.949delC p.L317Wfs*	c.913C>G p.P305A	c.1502G>A p.R501Q	c.483G>C p.E310D	c.11047A>G p.I3683V	c.10691G>A p.R3564H	c.7169C>G p.P2390R	c.635A>G p.Y218C	c.1448T>C p.L483P (p.L444P)	c.509G>T p.R170L (p.R131L)	c.635C>T p.S251L (p.S212L)	c.535G>C p.D179H (p.D140H)	c.1093G>A p.E365K (p.E326K)
rs number	rs113742727	Novel	Novel	rs112600292	rs61744200	rs72480423	rs115819951	rs116228685	rs148467516	rs149878646	rs421016	rs80356763	Novel	rs147138516	rs2230288
MAF	0.001039	-	-	0.00005304	0.003218	0.000137 9	0.00385 4	0.00027 27	0.00021 25	0.00139 7	0.001226	0.00000 3981	-	0.00012 76	0.01073
Patient ID															
10.783† (AAO: 13)		✔	✔												
88.28† (AAO:40)		✔	✔												
89.01† (AAO: 47)		✔	✔				✔								
94.69† (AAO: 30)		✔	✔	✔								✔			

11.861† (AAO: 63)						✔								
12.731† (AAO: 46)	✔				✔			✔	✔					
12.819† (AAO: 59)										✔			✔	
12.799+ (AAO: 49)											✔			
12.726+ (AAO: 43)													✔	✔
Scores														
SIFT	0 (D)	NA	NA	0 (D)	0.115 (T)	0.077 (T)	0.175 (T)	0 (D)	0.022 (D)	0.001 (D)	0.002 (D)	0.009 (D)	0.153 (T)	0.049 (D)	0.784 (T)
Polyphen-2	1 (D)	NA	NA	1 (D)	1 (D)	0.971 (D)	0.983 (D)	1 (D)	0.958 (D)	0.999 (D)	0.612 (D)	0.973 (D)	0.182 (B)	0.544 (D)	0.03 (B)
FATHMM	0.99207 (D)	NA	NA	-0.93 (T)	-0.14 (T)	-2.19 (D)	0.84 (T)	0.53 (T)	0.86 (T)	0.82 (T)	-5.92 (D)	0.88461 (D)	-5.55 (D)	-5.8 (D)	-5.77 (D)
MutationTaster	1 (D)	NA	NA	1 (D)	0.999987 (D)	0.593 (D)	1 (D)	1 (D)	0.999904 (D)	1.000 (D)	1 (D)	0.999945 (D)	0.744 (D)	0.632648 (T)	0.000175813 (T)
CADD	35	NA	NA	28.3	31	22.4	23.1	35	23.2	26.8	24.8	27.5	9.174	8.229	17.33

DANN	1.000 (D)	NA	NA	0.997 (D)	0.999 (D)	0.994 (D)	0.998 (D)	0.999 (D)	0.998 (D)	0.998 (D)	0.996 (D)	-	0.978 (T)	0.986 (T)	0.985 (T)
MetaSVM	-0.154 (T)	NA	NA	0.076 (D)	-0.7139 (T)	0.278 (D)	-0.636 (T)	-0.2818 (T)	-0.6621 (T)	-0.667 (T)	1.1096 (D)	1.0808 (D)	0.627 (D)	0.9028 (D)	0.6748 (D)
MetaLR	0.3525 (T)	NA	NA	0.545 (D)	0.1245 (T)	0.704 (D)	0.190 (T)	0.3699 (T)	0.2324 (T)	0.2 (T)	0.9738 (D)	0.978 (D)	0.813 (D)	0.9247 (D)	0.8499 (D)
LRT	0.000001 (D)	NA	NA	0 (D)	0 (D)	0.001115 (N)	0 (D)	0 (D)	0 (D)	0.000094 (N)	0.005056 (N)	0.000062 (D)	0.000438 (D)	0.007155 (N)	0.075111 (D)
PROVEAN	-3.66 (D)	-5.73 (D)	-3.83 (D)	-7.32 (D)	-0.67 (N)	-2.15 (N)	-0.486 (N)	-4.73 (D)	-4.21 (D)	-3.47 (D)	-5 (D)	-5.54 (D)	-2.42 (N)	-2.37 (N)	-1.2 (N)
GERP++	5.71	NA	NA	6.07	5.84	3.02	5.36	5.5	5.56	4.28	3.16	3.55	3.66	3.55	3.67
SIFT-Indel	NA	0.000 (D)	0.000 (D)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ACMG Classification	VUS (PP3 and PM2 apply)	VUS (PP3, PM2, and PM4 apply)	VUS (PP3, PM2, and PM4 apply)	Likely pathogenic (PP2, PP3, PM1, and PM2 apply)	VUS (PP3, PM1, and PM2 apply)	Likely pathogenic (PP2, PP3, PM1, and PM2 apply)	VUS (PP3 and PM2 apply)	VUS (PP3 and PM2 apply)	VUS (PP3 and PM2 apply)	VUS (PP3 and PM2 apply)	Pathogenic (PS3, PS4, PP2, PP3, PM1, and PM2 apply)	Likely pathogenic (PS3, PP2, PP3, and PM2 apply)	VUS (PP3 and PM2 apply)	Likely pathogenic (PS3, PS4, PP2, PP3, and PM2 apply)	Likely pathogenic (PS3, PS4, PP2, and PM1 apply)

D, deleterious; P, polymorphism; T, tolerated; B, benign; N, neutral; -,none; †,Positive family history; VUS, Variants of uncertain significance
>20 CADD score indicates that the variant is deleterious; > 3 GERP++ score indicates that the region where that variant is located is conserved

3.6 Validation of the prioritized variants and co-segregation analysis

The four *GBA* mutations previously reported as pathogenic, likely pathogenic or PD risk factors and the one novel *GBA* variant were validated using the method described in Part B of the study. This was done to determine whether the NGS approach was as effective as the nested PCR method for the detection of variants found only in *GBA* and not in its pseudogene. The p.L483P, p.R170L, p.D179H, pE365K variants were all validated in *GBA*, while the novel p.S251L variant was not found indicating that it might be occurring in the pseudogene. Thus, signifying that although the NGS is capable of detecting variants in the functional G*BA* gene validation with the specialised primers is still required. Furthermore, for those whom we had DNA of the affected family members the variants were screened to assess whether they segregate with the disease. We had DNA from the p.R170L carrier affected sister and the variant was found to be present in this individual. Thus, indicating that it is segregating with the disease in this family (Appendix 3). The p.D179H and p.E365K variants both occurred in one individual. Subsequently, they were screened in the carrier's affected siblings and nephew, their brother-in-law, their affected and unaffected cousins. The Sanger sequencing results revealed that the p.D179H is present in their affected nephew. The other non-*GBA* prioritized likely pathogenic variants were also selected for Sanger sequencing validation as well as screening in family members. The *PINK1* p.P305A variant was validated in the affected individual, who is also carrying the above mentioned *GBA* p.R170L. DNA of the carrier's affected sibling was assessed and their Sanger sequencing results indicated that the variant does not segregate with the disease in the family. The *PRKN* p.E310D variant was validated in the carrier and was assessed in their family members whom we had their DNA for analysis (Appendix 3). The variant was found to be present in their affected and unaffected daughters, but it was absent in the father and their unaffected son. This may indicate that the variant co-segregates with the disease in the family. To summarise the main findings obtained in Part A of the thesis, a flowchart of these results is provided in Figure 3.11.

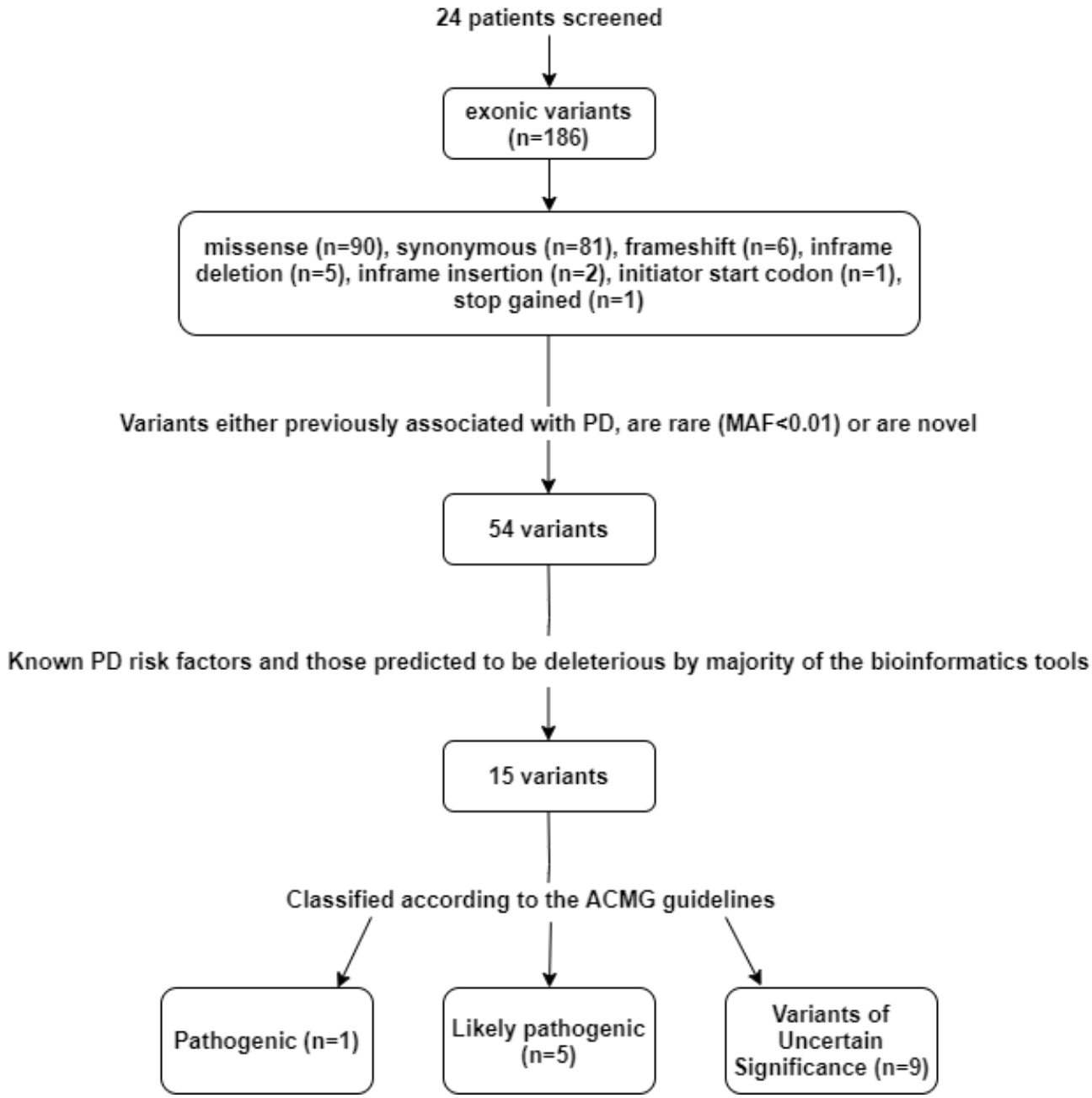


Figure 3.11 Flowchart summary of the results of Part A of the study. The 15 prioritized variants are shown in table 3.3. Only the clinically actionable (pathogenic and likely pathogenic) variants were selected for Sanger sequencing validation and screening in family members of the carriers.

Part B: Screening of *GBA* in Black South African patients

Additionally, for the second aim of this study, we set up a method to screen the *GBA* gene. Although this gene had been incorporated into the gene panel it is known that the presence of a highly homologous (96%) pseudogene (*GBAP1*) would complicate the analysis. Therefore, we used a nested PCR approach in which three large genomic fragments are first generated and then used as templates for *GBA*-specific exon amplification. We selected 30 PD patients of African ancestry for this analysis since for this gene there are no published reports of screening of this gene in African PD patients to date. This work was written up as a manuscript and submitted to the journal Neurobiology of Ageing where it is currently under review.

The submitted version of this manuscript is included in this section of the thesis results.

Mutation screening of the glucocerebrosidase (*GBA*) gene in South Africans of African ancestry with Parkinson's disease

Amokelani C. Mahungu [a], David G. Anderson [b], Anastasia C. Rossouw [c], Riaan van Coller [d], Jonathan A. Carr [e], Owen A. Ross [f,g], Soraya Bardien [a, *]

[a] Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[b] University of the Witwatersrand Donald Gordon Medical Centre, Neurology, University of the Witwatersrand, Johannesburg, South Africa,

[c] Division of Neurology, Department of Medicine, Faculty of Health Sciences, Walter Sisulu University, East London, South Africa

[d] Department of Neurology, School of Medicine, Faculty of Health Sciences, University of Pretoria, South Africa

[e] Division of Neurology, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

[f] Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, USA

[g] Department of Clinical Genomics, Mayo Clinic College of Medicine, Jacksonville, Florida, USA

* Corresponding author at: Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, PO Box 241, Cape Town, 8000, South Africa. Tel +27 21 938 9681

E-mail address: sbardien@sun.ac.za (S. Bardien)

Tables: 2

Figures: 0

Supplementary data: Table S1 and Table S2

Supplementary data: Figure S1

ABSTRACT

Mutations in glucocerebrosidase (*GBA*) are a major genetic risk factor for Parkinson's disease (PD) and display ethnic-dependent frequencies, e.g. mutations such as p.N370S and 84insGG are common in Ashkenazi Jewish patients. Notably, there are limited studies on Black patients from the African continent; hence, we conducted a study on 30 South African Black PD patients. All 11 exons of *GBA* were screened using a nested PCR approach to avoid pseudogene contamination. We identified previously described Gaucher's disease-associated variants, p.R120W in one patient [age-at-onset (AAO) of 35 years], and p.R131L in another patient (AAO 3O years) and in her affected sibling (AAO 45 years). Also, we found three previously-identified [p.K(-27)R, p.T36del, and p.Q497*], and two novel (p.F216L and p.G478R) variants. Screening of ethnic-matched controls for the novel variants revealed that the allele frequency of p.F216L was 9.9%, whereas p.G478R was not found in the controls. Studies such as these are important and necessary to reveal the genetic architecture underlying PD in the understudied patients of African ancestry.

Keywords: Glucocerebrosidase; *GBA* mutations; Parkinson's disease; South African; African ancestry

1. Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disorder resulting in various motor and non-motor related symptoms. Although the clinical and neuropathological features of PD have been extensively described, the etiology remains unclear. The *GBA* gene, which encodes the lysosomal enzyme glucocerebrosidase, has been consistently shown through various multicenter and genome-wide association studies to be a major risk factor for developing PD (Chang et al., 2017; Sidransky et al., 2009). *GBA* mutations are also known to cause Gaucher's disease (GD), an autosomal recessive lysosomal storage disorder, initially described by Phillipe Gaucher in the 1800s (Gaucher, 1882). *GBA* (OMIM 606463) is located on chromosome 1q22 and consists of 11 coding exons. Mutation screening of this gene is hampered by the presence of a highly homologous (97%) pseudogene (*GBAP1*), located 16kb downstream (Horowitz et al., 1989). This complication has led to many researchers conducting mutation screening of only selected regions of the gene.

GBA mutations are prevalent within Ashkenazi Jewish PD patients; in one study 19% of the 1,000 PD patients screened harbored mutations (Gan-Or et al., 2015). The most common pathogenic mutations in this population are p.N370S, p.R496H, 84insGG, IVS2+1G-A, pV394L, p.L444P and 370Rec (Gan-Or et al., 2015). These have been shown to be relatively rare in non-Jewish ethnicities (Zhang et al., 2018). It is hypothesized that diverse populations may harbor unique *GBA* mutations, which emphasizes the necessity for screening of the entire coding region in these individuals. *GBA*-associated PD risk in populations on the African continent have been understudied, with only three studies conducted thus far. One was in individuals of North African ancestry, mostly from Algeria (Lesage et al., 2011) and the other study was in North African Arab-Berber individuals from Tunisia (Nishioka et al., 2010). The third study was performed in South African individuals of European ancestry (Barkhuizen et al., 2017). This study identified known and putative pathogenic variants (including p.N370S and p.I368T), as well as risk factors in *GBA* (including p.E326K and p.T369M) which were found more commonly in patients than in controls (12.4% vs. 5.0%). To date, no studies have investigated *GBA* mutations in sub-Saharan African PD patients of non-European ancestry. Hence, the aim of the present study was to screen the entire coding region of *GBA* (all 11 exons) in a group of 30 South African Black PD patients to determine whether the mutations identified in other populations are also found in African populations.

2. Methods

Study participants: The study was approved by the Health Research Ethics Committee of Stellenbosch University, South Africa (2002/C059). All individuals provided written informed consent. Thirty Black South African PD patients (**Supplementary Table S1**) were recruited from the Movement Disorders clinic at Tygerberg Hospital in Cape Town as well as from various neurology clinics around the country. The mean age-at-onset (AAO) was 47.7 ± 8.4 years (range 30 - 57) and 53% were male. Two individuals reported a family history of the disorder: with both having an affected sibling and an affected parent (**Supplementary Table S1**). Participants were assessed by neurologists and met the UK Parkinson's Disease Society Brain Bank diagnostic criteria for PD diagnosis (W. R. G. Gibb and Lees, 1988). Parkinsonism was present in all patients, and no absolute exclusion criteria were present (Postuma et al., 2015).

Genetic analysis: For all 30 individuals, the 11 exons of *GBA* were PCR amplified using previously described primers (Stone et al., 2000). Firstly, the exons were amplified in three large fragments, consisting of exons 1-5 (2,972bp), exons 5-7 (2,049bp) and exons 8-11 (1,682bp) (**Supplementary Table S2**). This was followed by a nested PCR step to amplify individual exons from the three large fragments for each individual (Stone et al., 2000). Thereafter, Sanger sequencing was performed using the BigDye Terminator V3.1 sequencing kit (Applied Biosystems, San Francisco, CA, USA), following the manufacturer's protocol, at the Central Analytical Facilities (Stellenbosch, South Africa). The electropherograms were evaluated using BioEdit version 7.0.5.3 (Hall, 1999). The pathogenicity of non-synonymous variants was assessed with various *in-silico* tools (MutationTaster-2, CADD, polyphen-2, SIFT, MutPred, and FATHMM) using Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). Indels were assessed with SIFT-indel and PROVEAN (Choi & Chan, 2015; Hu & Ng, 2013). The Genome Aggregation Database (gnomAD) was used to assess the frequency of selected variants (Karczewski et al., 2019). In addition, the two novel variants were screened in ethnic-matched controls [n=110 individuals; average current age = 89.1 ± 5.2 years (range 81-105) and 48% were male].

3. Results

In total, five missense, one premature stop, one indel and one synonymous *GBA* variant were identified in 17 individuals (**Table 1**). All of the variants were heterozygous. We identified two novel [p.F216L (p.F255L) and

p.G478R (p.G517R); **Supplementary Figure S1**] as well as five known [p.K(-27)R (p.K13R), p.T36del (p.T75del), p.R120W (p.R159W), p.R131L (p.R170L), and p.Q497* (p.Q536*)] variants. The two novel variants were absent from the gnomAD, however, when we screened ethnic-matched non-PD controls, we found the allele frequency of p.F216L to be 9.9% (21/212 chromosomes). The p.G478R was not found in any the controls (0/220 chromosomes).

Notably, the p.R120W (p.R159W) pathogenic mutation was present in one individual. Also, a p.R131L (p.R170L) mutation, previously associated with GD, was found in one individual. This individual has an affected father and sister, and Sanger sequencing of the sister (AAO 45 years) revealed that she is also a carrier of this mutation. The *in-silico* pathogenicity predictions of the variants show that all except p.K(-27)R (p.K13R) are predicted to be deleterious (**Table 2**).

4. Discussion

Given the importance of *GBA* variants in PD risk, we set out to establish if Black South African patients harbored common mutations, such as those observed in the Ashkenazi Jewish PD patients. We did not find the common polymorphisms associated with PD risk [p.E326K (p.E365K) or p.T369M (p.T408M)]. Also, we did not identify any of the putative pathogenic variants identified in the previous study performed in South African individuals of European ancestry (Barkhuizen et al., 2017). We did, however, find the p.K(-27)R (p.K13R) variant in six individuals. In a previous study, this variant was found in two individuals with PD of North African Arabic ancestry, however, it was found to occur at similar frequencies (approximately 4%) in cases and controls (Nishioka et al., 2010). Another study reported the p.K(-27)R (p.K13R) variant in two individuals, who also harbored the *LRRK2* p.G2019S mutation, which is common in North African and Ashkenazi Jewish PD patients (Lesage et al., 2011). According to the gnomAD database, this variant is common in individuals of African ancestry (MAF = 0.07687). Five out of the six *in-silico* tools used predicted this variant to be benign (Table 2).

In addition, two individuals were found to harbor a three-base pair (TAC) in-frame deletion [p.T36del (p.T75del)] in exon 3 of *GBA*, which is reported in dbSNP as a variant of unknown pathogenicity. However, it was predicted to be damaging in our analysis using SIFT-indel and PROVEAN. Interestingly, p.T36del was found in 17/38 chromosomes (all heterozygous) from type 1 GD patients of African ancestry, thereby

suggesting that this allele may be relatively frequent in this population (Arndt et al., 2009) but further studies on its frequency are necessary. The other two known variants p.R120W [p.R159W] and p.R131L [p.R170L]) were identified in one individual each. p.R120W is a known cause of type 2 GD but is also found in type 1 GD cases, whereas p.R131L has been found in type 3 GD cases (Pastores and Hughes, 1993; DL. Stone et al., 2000; Wan et al., 2006). Both participants have young-onset PD (35 and 30 years) and one has a family history of PD. p.R120W (p.R159W) has been associated with an increased risk of developing PD, especially in non-Ashkenazi Jewish populations (Zhang et al., 2018). To our knowledge, p.R131L (p.R170L) has not previously been associated with PD. However, pathogenic *GBA* mutations have been shown to increase the risk of PD when in a heterozygous form. These two variants were predicted to be pathogenic by all of the *in-silico* tools (Table 2).

Interestingly, we identified two novel variants that are not reversions to the pseudogene. The p.F216L (p.F255L) variant was found in six individuals (with AAO's of 30, 45, 53, 55, 55 and 57 years) and it is predicted to be pathogenic by all six *in-silico* tools (**Table 2**). However, this variant was found in 9.9% of control chromosomes indicating that it is a common polymorphism. One of the individuals (AAO 53 years) is a compound heterozygote, carrying both the p.F216L (p.F255L) variant and a known stop gain variant p.Q497* (p.Q536*) which is predicted to be pathogenic by three of the *in-silico* tools (**Table 2**). p.Q497* (p.Q536*) is a change from glutamine at position 536 to a premature stop codon i.e. the protein is one amino acid shorter than the wild type. This has been reported in dbSNP as a variant of unknown pathogenicity. The other novel variant, p.G478R (p.G517R), was detected in one individual with an AAO of 37 years and was predicted to be pathogenic by all six *in-silico* tools (**Table 2**). It was not found in ethnic-matched controls and therefore further follow-up studies on this variant is necessary. Interestingly, the variant is at the same codon as a known pathogenic mutation (p.G478S [p.G517S]) previously found in a compound heterozygous state (other mutation was p.N370S) in a non-Jewish European type 1 GD patient (Beutler et al., 1993).

In conclusion, our study has shown that screening of the entire coding region of *GBA* in diverse populations is important, as unique variants may be identified. Our study is limited by a small sample size, and future studies on individuals of African ancestry should include larger sample sizes. Another limitation is that the majority of our study participants do not know if they have a family history of PD, limiting co-segregation

analysis. Despite these limitations, this study is important since it is the first report on *GBA*-associated PD risk in individuals of African ancestry. As new therapeutic targets for *GBA*-associated PD are being developed, including microRNAs and substrate reduction therapy, identification of *GBA*-mutation carriers who may benefit from participating in clinical trials, is essential (Ryan et al., 2019).

Acknowledgments

We thank the study participants for taking part in the study. The authors are financially supported by the National Research Foundation of South Africa (Grant Number: 106052), the South African Medical Research Council (Self-Initiated Research Grant), and Stellenbosch University. We also acknowledge the support of the NRF-DST Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town. Dr. Ross is funded by grants from the NIH, Michael J. Fox Foundation, the American Parkinson Disease Association (APDA) Mayo Clinic Information and Referral Center, Mayo Clinic APDA Center for Advanced Research and the Mayo Clinic Lewy Body Dementia Association (LBDA) Research Center of Excellence. Also, we acknowledge Oluwafemi Oluwole for providing assistance with the primers and Katelyn Cuttler for assistance with the screening of controls.

Conflict of interest

The authors declare no conflicts of interest.

Table 1 Heterozygous coding variants identified in *GBA* in Parkinson's disease patients of African ancestry

Participant ID no.	Exon	[1] Variant (traditional nomenclature)	[2] Variant (HGVS nomenclature)	cDNA position	Codon change	rs number (dbSNP)	[3] MAF from gnomAD (all)	MAF from gnomAD (African ancestry)	MAF from gnomAD (European ancestry; non-Finnish)
Nonsynonymous variants and indels									
43.59	2	p.K(-27)R	p.K13R	c.38A>G	A**A**G/A**G**G	rs150466109	0.007478	0.07687	0.0001629
12.170	2	p.K(-27)R	p.K13R						
10.313	2	p.K(-27)R	p.K13R						
10.314	2	p.K(-27)R	p.K13R						
11.895	2	p.K(-27)R	p.K13R						
11.910	2	p.K(-27)R	p.K13R						
11.835	3	p.T36del	p.T75del	c.222_224delTAC	GG**TAC**C/GG-C	rs761621516	0.00006367	0.0004409	0.000
11.962	3	p.T36del	p.T75del						
12.486	5	p.R120W	p.R159W	c.475C>T	**C**GG/**T**GG	rs439898	0.00002124	0.0001203	0.00001552
94.69	5	p.R131L	p.R170L	c.509G>T	C**G**C/C**T**T	rs80356763	0.000003981	0.000	0.000
60.39	7	p.F216L	p.F255L	c.765C>T	TT**C**/TT**A**	Novel			

84.52	7	p.F216L	p.F255L						
12.172	7	p.F216L	p.F255L						
12.177	7	p.F216L	p.F255L						
12.179	7	p.F216L	p.F255L						
12.178	7	p.F216L and	p.F255L and						
	11	p.Q497*	p.Q536*	c.1606C >T	**C**AG/**T**AG	rs758806595	0.00002510	0.0002309	0.000
96.87	11	p.G478R	p.G517R	c.1549G >C	**G**GC/**C**GC	Novel			
Synonymous variants									
84.52	10	p.Q432Q	p.Q471Q	c.1413G >A	CA**G**/CA**A**	rs12747811	No data		

[1] Amino acid numbering according to the traditional nomenclature used to name variants associated with Gaucher's Disease. This system uses the 40th amino acid as the first codon, as the first 39 amino acids are cleaved off in the active form of the protein (Hruska et al., 2008).
[2] Amino acid numbering according to the reference sequence NP_000148.2 and Human Genome Variation Society nomenclature (HGVS; http://varnomen.hgvs.org/).
[3] MAF (minor allele frequency) is taken from the Genome Aggregation Database of exome and genome sequences of various populations (gnomAD; https://gnomad.broadinstitute.org/).

Table 2 *In-silico* pathogenicity predictions of the variants identified in *GBA*

[1] **Variant**	**Scaled CADD v1.3 score (>15=deleterious)**	**MutPred score (>0.5=deleterious)**	**MutationTaster-2 prediction**	**FATHMM**	**PolyPhen-2 prediction (>0.5= deleterious)**	**SIFT Prediction (<0.05= deleterious)**
p.K13R	9.543	0.243	Polymorphism	Disease causing	Benign (0)	Tolerated (0.53)
[†] p.T75del						
p.R159W	28.6	0.827	Disease causing	Disease causing	Probably damaging (0.934)	Deleterious (0.02)
p.R170L	27.5	0.860	Disease causing	Disease causing	Possibly damaging (0.906)	Deleterious (0.01)
p.F255L	23.9	0.872	Disease causing	Disease causing	Probably damaging (1)	Deleterious (0)
p.G517R	24.8	0.679	Disease causing	Disease causing	Probably damaging (0.944)	Deleterious (0.03)
p.Q536*	36	Not scored	Disease causing	Disease causing	Not scored	Not scored

[1] Amino acid numbering according to the reference sequence NP_000148.2
[†] As this is an indel we used other tools such as SIFT-indel and PROVEAN to assess its pathogenicity

5. References

Arndt, S., Heitner, R., Lane, A., Ramsay, M., 2009. Glucocerebrosidase gene mutations in black South Africans with Gaucher disease. Blood Cells, Mol. Dis. 43, 129–133. https://doi.org/10.1016/J.BCMD.2009.02.008

Barkhuizen, M., Anderson, D.G., van der Westhuizen, F.H., Grobler, A.F., 2017. A molecular analysis of the GBA gene in Caucasian South Africans with Parkinson's disease. Mol. Genet. Genomic Med. 5, 147–156. https://doi.org/10.1002/mgg3.267

Beutler, E., Gelbart, T., West, C., 1993. Identification of Six New Gaucher Disease Mutations. Genomics 15, 203–205. https://doi.org/10.1006/geno.1993.1035

Chang, D., Nalls, M.A., Hallgrímsdóttir, I.B., Hunkapiller, J., van der Brug, M., Cai, F., Kerchner, G.A., Ayalon, G., Bingol, B., Sheng, M., Hinds, D., Behrens, T.W., Singleton, A.B., Bhangale, T.R., Graham, R.R., Bhangale, T.R., Graham, R.R., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. 49, 1511–1516. https://doi.org/10.1038/ng.3955

Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–7. https://doi.org/10.1093/bioinformatics/btv195

Gan-Or, Z., Amshalom, I., Kilarski, L.L., Bar-Shira, A., Gana-Weisz, M., Mirelman, A., Marder, K., Bressman, S., Giladi, N., Orr-Urtreger, A., 2015. Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology 84, 880. https://doi.org/10.1212/WNL.0000000000001315

Gaucher, P., 1882. Primary epithelioma of the spleen, idiopathic hypertrophy of the spleen without leukemia (doctoral thesis). Paris Fac. Med. Paris.

Gibb, W.R.G., Lees, A.J., 1988. A comparison of clinical and pathological features of young- and old-onset parkinson's disease. Neurology 38, 1402–1406. https://doi.org/10.1212/wnl.38.9.1402

Hall, T.A., 1999. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp. 41, 95–98.

Horowitz, M., Wilder, S., Horowitz, Z., Reiner, O., Gelbart, T., Beutler, E., 1989. The human glucocerebrosidase gene and pseudogene: structure and evolution. Genomics 4, 87–96.

Hu, J., Ng, P.C., 2013. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. PLoS One 8, e77940. https://doi.org/10.1371/journal.pone.0077940

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L.,

Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Consortium, T.G.A.D., Neale, B.M., Daly, M.J., MacArthur, D.G., 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv 531210. https://doi.org/10.1101/531210

Lesage, S., Condroyer, C., Hecham, N., Anheim, M., Belarbi, S., Lohman, E., Viallet, F., Pollak, P., Abada, M., Dürr, A., Tazir, M., Brice, A., French Parkinson Disease Genetic Group, F. the F.P.D.G., 2011. Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa. Neurology 76, 301–3. https://doi.org/10.1212/WNL.0b013e318207b01e

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. Genome Biol. 17, 122. https://doi.org/10.1186/s13059-016-0974-4

Nishioka, K., Vilariño-Güell, C., Cobb, S.A., Kachergus, J.M., Ross, O.A., Wider, C., Gibson, R.A., Hentati, F., Farrer, M.J., 2010. Glucocerebrosidase mutations are not a common risk factor for Parkinson disease in North Africa. Neurosci. Lett. 477, 57. https://doi.org/10.1016/J.NEULET.2009.11.066

Pastores, G.M., Hughes, D.A., 1993. Gaucher Disease, GeneReviews®. University of Washington, Seattle. https://doi.org/https://www.ncbi.nlm.nih.gov/books/NBK1269

Postuma, R.B., Berg, D., Stern, M., Poewe, W., Olanow, C.W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A.E., Halliday, G., Goetz, C.G., Gasser, T., Dubois, B., Chan, P., Bloem, B.R., Adler, C.H., Deuschl, G., 2015. MDS clinical diagnostic criteria for Parkinson's disease. Mov. Disord. 30, 1591–601. https://doi.org/10.1002/mds.26424

Ryan, E., Seehra, G., Sharma, P., Sidransky, E., 2019. GBA1-associated parkinsonism. Curr. Opin. Neurol. 1. https://doi.org/10.1097/WCO.0000000000000715

Sidransky, E., Nalls, M.A., Aasly, J.O., Aharon-Peretz, J., Annesi, G., Barbosa, E.R., Bar-Shira, A., Berg, D., Bras, J., Brice, A., Chen, C.-M., Clark, L.N., Condroyer, C., De Marco, E. V, Dürr, A., Eblan, M.J., Fahn, S., Farrer, M.J., Fung, H.-C., Gan-Or, Z., Gasser, T., Gershoni-Baruch, R., Giladi, N., Griffith, A., Gurevich, T., Januario, C., Kropp, P., Lang, A.E., Lee-Chen, G.-J., Lesage, S., Marder, K., Mata, I.F., Mirelman, A., Mitsui, J., Mizuta, I., Nicoletti, G., Oliveira, C., Ottman, R., Orr-Urtreger, A., Pereira, L. V, Quattrone, A., Rogaeva, E., Rolfs, A., Rosenbaum, H., Rozenberg, R., Samii, A., Samaddar, T., Schulte, C., Sharma, M., Singleton, A., Spitz, M., Tan, E.-K., Tayebi, N., Toda, T., Troiano, A.R., Tsuji, S., Wittstock, M., Wolfsberg, T.G., Wu, Y.-R., Zabetian, C.P., Zhao, Y., Ziegler, S.G., 2009. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med. 361, 1651–61. https://doi.org/10.1056/NEJMoa0901281

Stone, D., Tayebi, N., Orvisky, E., Stubblefield, B., Madike, V., Sidransky, E., 2000. Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease. Hum. Mutat. 15, 181–188. https://doi.org/10.1002/(SICI)1098-1004(200002)15:2<181::AID-HUMU7>3.0.CO;2-S

Wan, L., Hsu, C.M., Tsai, C.H., Lee, C.C., Hwu, W.L., Tsai, F.J., 2006. Mutation analysis of Gaucher disease patients in Taiwan: High prevalence of the RecNciI and L444P mutations. Blood Cells, Mol. Dis. 36, 422–425. https://doi.org/10.1016/j.bcmd.2006.02.001

Zhang, Y., Shu, L., Sun, Q., Zhou, X., Pan, H., Guo, J., Tang, B., 2018. Integrated Genetic Analysis of Racial Differences of Common GBA Variants in Parkinson's Disease: A Meta-Analysis. Front. Mol. Neurosci. 11, 43. https://doi.org/10.3389/fnmol.2018.00043

CHAPTER 4: DISCUSSION

In the present study, a custom-designed resequencing gene panel was successfully set up for the screening of 23 PD-causing genes in our local populations. Furthermore, recognizing the complications of screening the *GBA* gene using NGS methods, due to the presence of a highly homologous pseudogene, we also successfully set up a screening method in our laboratory, to identify and validate variants that occur in the functional *GBA* gene. With these techniques, we will be able to investigate the genetic factors that contribute to monogenic forms of PD in South African patients.

The major findings of part A of the study (the gene panel) included 15 variants that were prioritized and classified according to the ACMG guidelines. A number of these are clinically actionable, i.e. the *PINK1* p.P305A and the *PRKN* pE310D variants which, according to ACMG guidelines, are classified as likely pathogenic. Clinically significant *GBA* variants (p.L483P [p.L444P], p.R170L [p.R131L], p.E365K [p.E326K] and p.D179H [p.D140H]) were identified with the gene panel and validated by the nested PCR and Sanger sequencing method. In part B of the study (*GBA* screening of 30 PD patients of African ancestry, for which a manuscript was submitted), we identified eight variants, two of which are novel (p.F255L [p.F216L] and p.G517R [p.G478R]).

4.1 PD gene panel

The gene panel included 23 candidate genes that have been implicated as causal factors for monogenic forms of PD and were selected by performing an extensive literature search. The majority of the selected candidate genes have not previously been investigated in SSA populations (Williams et al., 2018). For the few that were investigated *(LRRK2, PRKN, SNCA, PINK1, DJ-1, ATP13A2, EIF4G1,* and *VPS35*), only the mutations shown to be disease-causing in European and Asian populations were examined. The known pathogenic mutations were not found to be common in PD cases of SSA ancestry. For instance, in a study on Nigerian PD cases, mutations in the *LRRK2, PRKN,* and *ATXN3* genes were not frequently found (Okubadejo et al., 2008). Another study found variants in the *EIF4G1* gene not to be common in control samples from the Central African Republic, Democratic Republic of Congo, Nigeria, Namibia, Senegal and South Africa (Tucci et al., 2012). Additionally, in a Zambian study one novel *LRRK2* variant p.A1464G and two likely pathogenic *PRKN* exon 2 and 4 deletions were reported (Yonova-Doing et al., 2012). While another study in Ghanaian PD patients that investigated exons 31 and 41 of *LRRK2* did not identify any variants (Cilia et al., 2012). Compared to other SSA countries, most PD genetic studies (eight) have been conducted in South Africa (Williams et al., 2018). These studies only examined disease-causing variants in the *PRKN, LRRK2, SNCA, PINK1, DJ-1, UCH-L1, ATP13A2, MAPT, SNCAIP, EIF4G1,* and the *VPS35* genes. The studies reported pathogenic variants in *PRKN* (p.H200Q, D280N, p.E310D, p.R402C, exon deletions, and duplications), *PINK1* (p.P305A and p.Y258X), *SNCA* (whole gene triplication) and *LRRK2* (p.G2019S) in less than 20 patients (Bardien et al., 2009; Bardien et al.,

2010; Keyser et al., 2010a, 2010b). Thus, it is hypothesized that novel variants in the known PD genes may be causal factors in SSA PD cases as the common mutations have not been frequently encountered. This supports the importance of screening the entire coding regions of all the known PD genes to identify variants that might be a unique cause of monogenic forms of PD in these populations. The studies mentioned above mostly used approaches such as Sanger sequencing, MLPA, and HRM which only allow screening of a single gene or variant at a time. Thus, a robust high-throughput method such as an NGS gene panel would be more beneficial as it allows screening of multiple samples for variants in all regions of interest simultaneously, and at a lower cost when considering the number of genes screened. Furthermore, once implemented this method can also be applied to screen PD cases from other SSA countries through collaborations.

Previously, our research group used the commercially-available Ion AmpliSeq™ Neurological Research Panel to screen for PD causing mutations in our local populations (Oluwole, 2019). This panel included 751 genes associated with neurological diseases as well as brain function, of which only 12 were PD genes *(ATP13A2, DJ-1, EIF4G1, FBXO7, GIGYF2, HTRA2, LRRK2, PINK1, PLA2G6, PRKN, SNCA, VPS35)*. This panel has now been discontinued. Thus, it was imperative for us to design a new gene panel that would also offer us sufficient coverage of all coding regions. For our study, we obtained over 200 x-fold depth of coverage for both our sequencing runs (Table 3.1 and 3.2). A minimum of 30 x-fold depth of coverage was required. Notably, the Agilent's SureDesign web-based application for designing the gene panel was relatively easy to use and the gene targets were selected by searching the gene names and locations from Ensembl's database. Furthermore, we viewed the regions on the target genes that the designed probes would hypothetically cover. Thus, this application can be easily used to add or remove any PD genes, as informed by the literature.

Custom-designed targeted gene panels have been widely used to screen for disease-causing mutations in PD genes in cases with an unknown genetic cause. For instance, one study developed a movement disorder gene panel that included 20 PD genes (*PINK1, PRKN, LRRK2, PARK7, SNCA, FBXO7, SYNJ1, DNAJC6, VPS35, EIF4G1, DNAJC13, UCHL1, GIGYF2, HTRA2, GBA, MAPT, GCH1, PLA2G6, GRN,* and *TMEM230*) and detected pathogenic variants in 16.9% of their study population (Reale et al., 2018). Recently, another study developed a PD gene panel that included 15 of the PD genes (*SNCA, LRRK2, PARK2, PINK1, PARK7, GIGYF2, ATP13A2, UCHL1, PLA2G6, FBXO7, EIF4G1, VPS35, ACMSD, APOE,* and *GBA*), and they reported five novel pathogenic variants (Gorostidi et al., 2016). In our study, we identified putative pathogenic variants in 37.5% (9 of the 24) of the patients studied. Although our sample size was small compared to other studies, we detected more putative pathogenic variants. This may be accounted for by the fact that we screened more PD genes. For instance, some of our prioritized variants were found in *DNAJC13, VPS13C* and *RIC3* genes which were not included for screening by the previous studies.

The sensitivity of the custom-designed gene panel was assessed by including positive controls, with previously validated PD-causing mutations. We obtained 100% sensitivity for the two SNV's (*PRKN* p.G430D

and *LRRK2* p.G2019S) and the short indel (40 bp deletion *in PRKN* resulting in a frameshift [p.P113Tfs*]) with sufficient coverage (>30x). However, we did not identify any of the CNV's (*SNCA* triplication, and the *PRKN* exon 3 and exon 4 deletions) in our positive controls with the variant discovery pipeline that was employed. This suggests that a CNV specific pipeline is required for the identification of large rearrangements. Multiple algorithms for CNV analysis have been developed (Ellingford et al., 2017), however, they are not as well established as algorithms for detecting SNV's. Nevertheless, it is anticipated that as NGS technologies advance, robust bioinformatics tools will be developed for analysing all types of genetic variations including structural changes.

Although our gene panel was successfully designed, and sufficient coverage on our target regions was obtained, we also experienced so-called 'sequence wastage'. Many of our sequencing reads mapped in off-target regions. We hypothesize that the low on-target percentage may have resulted from inadequate amount of library, use of samples with poor quality DNA, the presence of EDTA or salts in some of our samples, and the occurrence of pseudogenes. Firstly, the electropherogram showing the fragment size distribution of the pooled library showed a low fluorescence signal intensity (Figure 3.5). This may indicate that the concentration of the pooled library was not optimal. Thus, we recommend the use of automated liquid handlers to automate library construction and improve the quality of the libraries produced (Kong et al., 2015). Secondly, it has been indicated that the quality of the DNA samples used for NGS may also result in inadequate amounts of library (Kapa biosystems, 2014). Although we tried to only include samples with high-quality genomic DNA, some of our sample's quality was not adequate for NGS application. Thirdly, few of our DNA samples were found to contain EDTA or salts, which may hamper NGS library construction. Thus, we attempted to use recently extracted samples as our new DNA extraction kit (Macherey-Nagel) may have less EDTA compared to the old method. However, we found that the new DNA extraction method buffer also contains high amounts of EDTA or salts even though it was not indicated by the supplier. Thus, it is recommended that DNA extraction methods that utilize fewer salts and EDTA are employed by our research group for extracting genomic DNA for NGS purposes. Alternatively, $MgCl_2$ can be used to remove excess EDTA (Kapa biosystems, 2014). Interestingly, we observed that prior to pooling, our libraries had varying amounts of DNA (between 279 and 750 ng). We hypothesize that although calculations were made to pool the libraries in equimolar amounts, there was no uniformity in the libraries before pooling. Thus, we recommend the implementation of methods such as spectrophotometric or fluorometric assays to normalize the DNA concentration and quantity, to ensure that all libraries are uniform, prior to pooling. Additionally, uniformity of the libraries can also be improved by building distinct protocols depending on the amounts and quantity of each DNA sample. Alternatively, adapter blocking oligos can be added to the library construction protocol as they have been shown to improve target capture for In-Solution Hybridization-based capture methods (Brookman-Amissah, 2014).

Initially, we hypothesized that the high off-target percentage might be a consequence of 'over sequencing' as we used a full Ion PI™ Chip Kit v3 for the first run for sequencing of 16 samples. The chip is expected to produce about 60-80 million reads which are a lot of reads for only 16 samples. Thus, for the second run, we only used half of the chip. However, this did not make a difference since the on-target percentage was similar to the first runs. In future, we recommend that the required coverage should be calculated prior to sequencing to determine how much of the chip should be used and to assess if it will improve the on-target percentage. Additionally, we hypothesized that the presence of pseudogenes or related genes whose sequence is highly similar to our target genes may have also contributed to the high off-target percentage. For instance, some of the reads were found to map to *GBAP1, GBA's* highly homologous pseudogene, and other reads aligned to *VPS35's* pseudogene *(VPS35P1)*. The off-target mapping that may have resulted from the presence of pseudogenes is difficult to overcome. However, one study demonstrated that a *GBA* specific NGS screening method can overcome the pseudogene complications by using *GBA* specific primers to amplify the gene before sample preparation and using long-range PCR for the library construction (Zampieri et al., 2017). Notably, it has been indicated that some level of off-target sequencing is unavoidable, and may result from probe-panel specific uninhibited hybridization (Hasin-Brumshtein et al., 2018). It is imperative that all these issues are considered for future optimization of the custom gene panel.

Overall, from the sequencing data of the 24 patients, exonic variants were identified in all our gene targets except for *SNCA, CHCHD2,* and *GCH1*. *SNCA* mutations have been indicated to be a rare cause of familial PD (Kasten and Klein, 2013). Thus, it is expected that variants in *SNCA* will not be as common. Mutations in *CHCHD2* and *GCH1* are thought to be a rare cause of PD, thus further studies are needed in diverse populations (Lunati et al., 2018). Notably, the majority of the exonic variants occurred in *LRRK2* and *VPS13C*. These genes are large, and they are therefore expected to harbour more variants. *LRRK2* contains 51 exons coding for 2527 amino acids while *VPS13C* consists of 86 exons encoding 3753 amino acids.

In total, from the variants that were detected in our study cohort, only 15 variants in 9 individuals were prioritised for further study (Table 3.3). These variants and the individuals they were found in, will be discussed in the following section.

Individual **10.783** is a male PD patient of Afrikaner ethnicity with an AAO of 13 years, and a family history of PD exhibiting an AR inheritance pattern (Appendix 3; Family ZA_450). The individual's affected sister also has an AAO of 13 years. Two novel *FBXO7* frameshift variants p.L317Gfs (c.948_949delTC) and p.L317Wfs* (c.949delC) were identified in this patient (Table 3.3). Both variants were classified as VUS (ACMG criteria) based on the following evidence:

- The variants are predicted to be deleterious by both SIFT-indel and PROVEAN (PP3 applied).
- They are absent from the public databases (PM2 applied).
- Both variants introduce multiple premature stops (PM4 applied).

Notably, the variants seem to occur at the same protein position. However, the p.L317Gfs is a two-base deletion of both the T at cDNA position 948 and a C at cDNA position 949. On the other hand, the p.L317Wfs* variant is a deletion of the C at cDNA position 949. *FBXO7* variants are typically associated with AR early or juvenile-onset PD (Joseph et al., 2018) which fits with the juvenile-onset in this patient. Nevertheless, Sanger sequencing validation of these novel frameshift variants is necessary to assess if they are not false positives as Ion torrent sequencing is known to have a higher chance of detecting indels related to homopolymer areas (Loman et al., 2012).

Interestingly, both variants were also identified in three additional patients in our study cohort (Table 3.3). Individual **88.28** is a male patient of Mixed ancestry with an AAO of 40 years, and a family history that exhibits a possible X-linked or AR inheritance pattern (Appendix 3; Family ZA_228). Notably, pathogenic variants in the X-linked *RAB39B* gene were not detected in this proband. The individual's grandfather is recorded as having PD however, his DNA was not available for analysis. This individual was found to be heterozygous for the two variants.

Individual **89.01** is a European female patient with an AAO of 47 years, and a family history with possible AD inheritance (Appendix 3; Family ZA_236). The patient's father is recorded as having PD however their DNA was not available. This individual is heterozygous for a *VPS13C* p.I3683V (rs115819951) variant as well as the two novel *FBXO7* frameshift variants (Table 3.3). The *VPS13C* p.I3683V was classified as a VUS (ACMG criteria) based on the following evidence:

- The variant was predicted to be likely deleterious by six of the *in-silico* tools (PP3 applied).
- It has a MAF 0.003854, demonstrating that it is rare (PM2 applied).

Interestingly, p.I3683V was reported in 1089 alleles out of 282,536 of which 765 were from non-Finnish European populations (https://gnomad.broadinstitute.org/variant/15-62146742-T-C). This may indicate that the variant is common within this population. However, screening in additional PD cases is required to determine this. It is plausible that the *VPS13C* variant and the two *FBXO7* variants acting together may increase the risk of PD in this individual. Functional studies on the effect of these variants on their respective proteins are required to determine their pathogenicity.

Furthermore, another individual (**94.69**) who has the two *FBX07* frameshifts is also a carrier of the pathogenic *GBA* p.R170L (p.R131L) as well as the *PINK1* p.P305A variant (Table 3.3). Individual **94.69** is a female patient of African ethnicity with an AAO of 30 years, and a family history resembling an AD inheritance pattern (Appendix 3; Family ZA_8). The individual has an affected sister whose DNA was available for analysis. Sanger sequencing confirmed the presence of two heterozygous variants, *PINK1* p.P305A and *GBA* p.R170L. Both variants were also assessed in the affected sister with Sanger sequencing. The *PINK1* variant was not present in the sister. This variant was assigned a likely pathogenic status (ACMG criteria) based on the following evidence:

- The variant occurs in a gene (*PINK1*) where most of the pathogenic variants are missense (PP2 applied).
- Ten of the *in-silico* tools predicted it to be deleterious (PP3 applied).
- It is in an evolutionarily conserved region, the kinase domain which is essential for kinase activity and selectivity of substrates. This is also the genes mutational hotspot (Corti et al., 2011) (PM1 applied).
- It is reported in gnomAD with a MAF of 0.00005304 (rs112600292) (PM2 applied).

Interestingly, the variant had previously been detected in the same individual by our research group and was found to be absent in the affected sister (Keyser et al., 2010a). In that study, the variant had been screened in 108 controls and it was discovered in 2 individuals. Owing to its location in the conserved kinase domain it was postulated that the variant might disrupt the autophosphorylation capability of *PINK1*. However, since it was not identified in the affected sibling and it was found in 1.9% of the controls the study suggested that the variant might be a polymorphism. Although the pathogenicity of this variant is not fully elucidated, previous studies have indicated that deleterious heterozygous *PINK1* mutations are disease modifiers or may increase the risk of PD or especially in familial PD cases (Abou-Sleiman et al., 2006; Klein et al., 2007).

Interestingly, the *GBA* p.R170L variant was detected in individual **94.69**'s affected sister and thus possibly may co-segregate with the disease in the family. This variant is a well-established cause of type 2 and 3 GD (Pastores and Hughes, 1993). We assigned a likely pathogenic (ACMG criteria) classification based on the following evidence:

- The variant has been shown to decrease the *GBA* enzyme (beta-glucocerebrosidase) activity (Stone et al., 2000) (PS3 applied).
- It is found to occur in a gene where most of the pathogenic variants are missense (PP2 applied).
- It was predicted to be deleterious by 10 of the *in-silico* tools (PP3 applied).
- Finally, it is reported with a MAF of 0.000003981 (rs80356763), indicating that it is not common (PM2 applied).

Although the variant has not been previously associated with the risk of developing PD, heterozygous pathogenic or severe GD mutations have been implicated to contribute to PD susceptibility (Gan-Or et al., 2015). Thus, indicating that this variant might be a risk factor in this individual and the affected sister who is also a carrier. The variant was also identified with the *GBA* Sanger sequencing method and it is described in the prepared manuscript. Given that the individual exhibits an AD inheritance pattern, it more likely that the *GBA* variant is a causal factor. Furthermore, all 11 exons of *GBA* were screened in this individual and they were not found to carry any of the known pathogenic PD mutations besides the *PINK1* variant. However, the activity of the beta-glucocerebrosidase enzyme will have to be assessed in the carriers to evaluate the activity and levels of the *GBA* enzyme.

Individual **11.861** is a European Afrikaner female patient with an AAO of 63 years, and a family history that exhibits an AD inheritance pattern (Appendix 3; Family ZA_459). The individual has a daughter diagnosed with PD and DNA of her affected daughter and three unaffected family members was available for genetic screening. This individual is a heterozygous carrier of a *PRKN* p.E310D (rs72480423) variant (Table 3.3). The variant was validated with Sanger sequencing in the proband and it was also present in a heterozygous state in her affected and unaffected daughters. The unaffected carrier is now 47 years, and her mother and sister were diagnosed with PD at ages 63 and 49 years, respectively. Thus, we suggest that that the unaffected sister is followed up to determine if she will later develop PD. This variant was assigned a likely pathogenic classification (ACMG criteria) based on the following evidence:

- *PRKN* has low rates of benign missense mutations and most of the disease-causing variants that occur within this gene are missense (PP2 applied).
- It is predicted to be deleterious by eight *in-silico* tools (PP3 applied).
- The variant is in a conserved region of the gene that has functional importance (between the RING1 and IBR domains) (PM1 applied).
- It has a MAF of 0.0001379, indicating that it is rare (PM2 applied).

Although homozygous or compound heterozygous *PRKN* mutations are known to cause PD, heterozygous mutations are implicated to increase the risk of PD and result in an age of onset variability (Klein et al., 2007). Thus, we hypothesize that the p.E310D variant might be increasing the risk of PD in this family. Interestingly, the variant was previously identified in one European Afrikaner individual with an AAO of 42 presenting with typical PD by our research group (Bardien et al., 2009). Subsequently, it was found to be absent in 110 ethnically matched controls. The variant was also reported in 14 Polish PD patients and suggested to be a polymorphism within this population (Oczkowska et al., 2015). That study proposed that the variant may result in incomplete penetrance and that it may lead to preclinical changes in the brain and increase the risk of PD. Although this variant is currently reported as a VUS on ClinVar, this evidence warrants its reclassification as a likely pathogenic variant or PD risk factor.

Individual **12.731** is a male patient of Mixed ancestry with an AAO of 46 years and exhibits an AR inheritance pattern (Appendix 3; Family ZA_538). The patient has an affected cousin whose DNA is not available for genetic screening. This individual was found to harbour four heterozygous variants (Table 3.3): *PINK1* p.R501Q, *DNAJC1*3 p.R1877Q, and p.R3564H and p.P2390R both in *VPS13C*, all of which were classified as VUS (ACMG criteria) based on the following evidence:

- The *PINK1* p.R501Q (rs61744200) variant was predicted to be deleterious by six of the *in-silico* tools (PP3 applied), it is located in the conserved kinase domain (mutation hotspot) (PM1 applied) (Corti et al., 2011) and it has a MAF of 0.003218 (PM2 applied). Interestingly, the variant has an allele frequency of 0.03251 from the population data of Africans/African Americans recorded on gnomAD

(https://gnomad.broadinstitute.org/variant/1-20976940-G-A). Thus, indicating that the variant is likely a polymorphism in African/American populations.

- The *DNAJC13* p.R1877Q (rs113742727) variant was predicted to be deleterious by nine of the *in-silico* tools (PP3 applied) and has a MAF of 0.001039 (PM2 applied). Interestingly, the variant was found to have an allele frequency of 0.01078 from population data of Africans/African Americans (https://gnomad.broadinstitute.org/variant/3-132235602-G-A). This may also indicate that it is a polymorphism in African/African American populations.
- The *VPS13C* p.R3564H (rs116228685) and p.P2390R (rs148467516) variants were each predicted to be deleterious by seven of the *in-silico* tools (PP3 applied) and had MAF of 0.0002727 and 0.0002125 respectively (PM2 applied). Interestingly, the p.R3564H variant was detected in 77 alleles out of 282,404 of which 71 were from non-Finnish Europeans (https://gnomad.broadinstitute.org/variant/15-62160901-C-T) and the p.P2390R variant was found in 60 alleles from African populations out of 282,412 (https://gnomad.broadinstitute.org/variant/15-62212445-G-C). This may also show that these variants are common within these populations.

Notably, it is recognized that all of the variants identified in this individual are in the ARPD genes and this individual exhibits an AR inheritance pattern. This warrant further investigation of these variants in larger numbers of PD cases. Furthermore, screening of the variants that were detected in South African patients belonging to the unique ethnolinguistic groups in ethnically matched controls is essential. As the African populations are diverse, they are not well represented on gnomAD and are understudied. Furthermore, the Mixed ancestry South African populations are highly admixed. This warrants screening of more cases and controls within these populations.

Individual **12.819** is a female patient of African ethnicity with an AAO of 59 years and exhibits an AR inheritance (Appendix 3; Family ZA_476). The person's paternal cousin is recorded as having PD; however, their DNA was not available for analysis. Two heterozygous variants were detected in this individual, in *GBA* (p.S251L [p.S212L]) and *RIC3* (p.Y218C) (Table 3.3). The *GBA* p.S251L variant is novel, it was only identified in this individual and was not present in the human population allele frequency databases. Sanger sequencing results of the variant revealed that it is not present in the functional gene. Thus, indicating that the variant may be in the pseudogene *(GBAP1)*. This emphasizes the importance of validating all *GBA* variants that were detected with the gene panel using the specialized primers. Nonetheless, we assigned a VUS classification (ACMG criteria) for this variant (ACMG criteria) based on the following evidence:

- It is predicted to be deleterious by six of the *in-silico* tools (PP3 applied).
- It was found in the population databases (PM2 applied).

The *RIC3* p.Y218C variant was also assigned a VUS classification (ACMG criteria), based on the following evidence:

- It was predicted to be deleterious by seven of the *in-silico* tools (PP3 applied).
- It is reported with a MAF of 0.001397 (rs149878646) indicating that it is not common (PM2 applied).

Mutations in *RIC3* are associated with ADPD and this individual's pedigree indicates an AR inheritance pattern. Making it questionable whether this variant is a causal factor. However, further investigation of the variant pathogenicity, screening of African control subject and functional studies are required to elucidate the potential impact of this variant. Interestingly, the variant was reported to have an allele frequency of 0.01470 in African/African American individuals (https://gnomad.broadinstitute.org/variant/11-8148223-T-C). The variant allele frequency is greater than 0.01 thus indicating that the variant might be a polymorphism in African/African American populations. However, it is also noted that the African populations on gnomAD are not representative of the unique South African ethnolinguistic groups.

Individual **12.799**, is a European Afrikaner female patient with an AAO of 49 years and exhibits an AD inheritance pattern (Appendix 3; Family ZA_544). The patient's affected father is deceased, and their DNA was not available at the time of the study. This individual was found to be heterozygous for the *GBA* p.L483P (p.L444P; rs421016) (Table 3.3) variant which was validated by Sanger sequencing. This variant has been previously classified as a pathogenic variant for GD and a risk factor for PD and dementia on the ClinVar database. We classified the variant as pathogenic (ACMG criteria) based on the following evidence:

- Functional studies have indicated that the presence of the variant results in unstable beta-glucocerebrosidase and reduces the enzyme activity (Dvir et al., 2003). Furthermore, the variant has been shown to maintain only about 5-14% of the enzyme residual activity (active enzyme in solution) (Malini et al., 2014; Montfort et al., 2004) (PS3 applied).
- The prevalence of the variant has been reported to be significantly higher in PD cases compared to controls (Malek et al., 2018; Sidransky and Lopez, 2012) (PS4 applied).
- Most of the disease-causing mutations that occur in *GBA* are missense (PP2 applied).
- The variant was predicted to be deleterious by 10 of the *in-silico tools* (PP3 applied).
- It occurs in exon 10, and exons 8-10 of *GBA* are considered as the gene's mutational hotspots (PM1 applied).
- The variant has a MAF of 0.001226, indicating that it is not common (PM2 applied).

This variant is one of the widely studied *GBA* mutations implicated to increase the risk of PD (Zhang et al., 2018; Malek et al., 2018). Notably, it was shown to increase PD risk in non-Ashkenazi Jewish populations (Zhang et al., 2018). Although the mechanisms in which *GBA* mutations result in PD are not fully understood it is hypothesized that the underlying PD pathways such as the aggregation of alpha-synuclein, mitochondrial dysfunction, impairment of lysosomal and autophagy pathways are implicated (Schapira and Tolosa, 2010). One study investigating the p.L483P heterozygous variant in an MPTP mouse model of PD showed that the reduction of the beta-glucocerebrosidase enzyme that was facilitated by the variant and the resulting alpha-

synuclein aggregation, causes the dopaminergic neurons to be more vulnerable to MPTP toxicity (Yun et al., 2018). This further supports the involvement of the p.L483P variant in PD risk. Although not all *GBA* mutation carriers develop PD, this individual was found to not carry any other pathogenic mutations in the common PD genes. Thus, it is possible that the *GBA* variant might be a causal factor. However, the activity and levels of the beta-glucocerebrosidase enzyme will need to be measured to confirm the expected reduced *GBA* activity and the patient's fibroblast should also be collected to perform functional studies.

Lastly, the other two *GBA* variants (p.E365K [p.E326K] and p.D179H [p.D140H]) were identified in individual **12.726** (Table 3.3). The affected carrier is an Afrikaner male with an AAO of 43 years and exhibits an AD inheritance pattern (Appendix 3; Family ZA_253). The DNA of the individual's affected siblings, nephew, and cousin and three unaffected relatives was tested. The presence of both two variants in the patient was validated by Sanger sequencing. Interestingly, the p.D179H variant was detected in the individual's affected nephew (92.32/13.558) who has an AAO of 37 years (Appendix3). The p.E365K variant (p.E326K; rs2230288) is classified as a mild GD mutation as it results in the disease when it occurs with a severe GD mutation like p.L483P (p.L444P) (Montfort et al., 2004). Multiple studies have indicated that although the p.E365K variant does not cause GD it increases susceptibility to PD (Duran et al., 2013; Huang et al., 2018; Malek et al., 2018). Recent metanalysis studies have shown that the p.E365K variant is associated with increased PD risk in Asian and European populations (Zhang et al., 2018; Huang et al., 2018). Interestingly, the variant was found to occur in 12% of PD cases compared to 5% of the controls in a study of 110 South African PD patients of European ancestry (Barkhuizen et al., 2017). This variant was classified as likely pathogenic (ACMG criteria) according to the following evidence:

- Functional studies have shown that the variant alone results in a mild enzyme activity reduction and resulted in about 25% enzyme residual activity when it occurs with a combined allele (Horowitz et al., 2011; Malini et al., 2014) (PS3 applied).
- The prevalence of the variant in PD cases has been reported to be higher than in control populations (Huang et al., 2018; Meeus et al., 2012; Nalls et al., 2013) (PS4 applied).
- The variant occurs in a gene where the disease-causing mutations are predominantly missense (PP2 applied).
- Furthermore, it is located in exon 8 of *GBA* which is a mutational hotspot (PM1 applied).

The other *GBA* variant, p.D179H (p.D140H; rs147138516), identified in this individual and their nephew is reported as a pathogenic variant for GD on ClinVar database. Interestingly, both variants (p.E365K and p.D179H) were previously identified as a complex allele in an individual of Irish/Polish and Mexican ancestry diagnosed with mild type 1 GD (Eyal et al., 1991; Grace et al., 1999). His genotype was p.D179H +p.E365K /p.K196Q [p.K157Q]). The complex allele (p.E365K and p.D179H) was also identified in a French PD patient

and found to be absent in 391 controls (Lesage et al., 2011). Another study detected the p.D179H variant in a PD patient of Belgium descent with dementia (Meeus et al., 2012). Thus, it is implicated as a PD and dementia risk factor (Crosiers et al., 2016; Meeus et al., 2012). It was concluded that the variant follows a non-Mendelian inheritance in this family as it was not identified in the proband's sister who is the mother of the individual who is also a carrier (the nephew) (Appendix 3; Family ZA_253). DNA of the proband's sister was screened twice to further support this conclusion. We also suspect the occurrence of a sample swap; this will be determined once a rebleed of the patient is obtained. The variant was classified as likely pathogenic (ACMG criteria) according to the following evidence:

- The variant was indicated to reduce the enzyme activity by functional studies (Santamaria et al., 2008) (PS3 applied).
- It has been reported in more PD cases vs controls by previous studies (Lesage et al., 2011; Nalls et al., 2013; Crosiers et al., 2016; Malek et al., 2018) (PS4 applied).
- Subsequently, the variant occurs in a gene where putative mutations are mostly missense (PP2 applied).
- It was predicted to be deleterious by six of the pathogenicity scoring *in-silico* tools (PP3 applied).
- Furthermore, it is reported as a rare variant with a MAF of 0.0001276 (PM2 applied).

Considering the strong Mendelian inheritance shown by the family's pedigree (Appendix 3; Family ZA_253). We hypothesize that the two variants are likely mild risk factors or disease modifiers. Furthermore, we suspect that they might be other causal factors contributing to PD in this family that was not in our target regions, were filtered out by the selection criteria or were not detected with the technology (e.g. deep intronic variants).

4.2 *GBA* screening

Although severe and mild *GBA* mutations have been shown to increase the risk of PD, to our knowledge there has not been a *GBA* study on PD patients of African ancestry. Thus, for the second part of the study, we aimed to establish a method for *GBA* screening that we can use to investigate the entire coding region for any *GBA* genetic variants that may have clinical significance in the pathogenesis of PD in our African PD patients. We identified six putative variants that were predicted to be deleterious by multiple *in-silico* tools. Furthermore, two of the pathogenic variants were novel (p.F255L [p.F216L] and p.G517R [p.G478R]). These were subsequently screened in ethnically matched controls and found to occur at a frequency of 9.9% and 0%, respectively. Interestingly, the novel p.G517R [p.G478R] variant is located at the same codon where a previous pathogenic GD disease mutation (p.G517S [p.G478S]) occurs (Beutler, 1993a; Beutler et al., 1993b). One of the pathogenic variants (p.R159W [p.R120W]) we detected is known to cause GD and it has been indicated to increase the risk of PD in non-Ashkenazi Jewish populations (Zhang et al., 2018). The last three

variants identified include the pathogenic p.R170L (p.R131L) variant described in individual **94.69**, a deleterious frameshift deletion p.T75del (p.T36del) indicated to be common in African populations and a stop gained variant of unknown pathogenicity (p.Q497* [p.Q536*]) (Arndt et al., 2009). Notably, although the contributions of these variants to PD are not fully established, studies have shown that heterozygous pathogenic and 'mildly deleterious' GD mutations increase the PD risk in multiple populations worldwide (Gan-Or et al., 2015; Zhang et al., 2018).

4.3 Study limitations

Although we were successful in setting up the targeted gene panel and the *GBA* screening method, our study had several limitations. Firstly, we encountered sequencing wastage as the majority of our sequencing reads (~70%) were mapping to off-target regions. However, we had sufficient coverage on our target regions to accurately call variants. Secondly, we were unable to detect CNV's in our positive controls. Thirdly, for our study, we designed our gene panel to predominantly screen the exonic regions of our target genes and approximately 10 bp into the introns. Although exonic variants are more likely to be disease-associated, intronic variations located further into the intron that may impair splicing and result in aberrant mRNA, may not be identified with our method (Vaz-Drago et al., 2017). Fourthly, the prioritization of variants was solely based on filtering strategies and criteria employed by previous studies on PD, this may result in incorrect ranking or skipping of some of the likely pathogenic variants. For example, we utilized a cut off MAF of 0.01 for rare variants, although some PD studies have recommended a MAF of 0.03 instead of 0.01 (Robak et al., 2017). The less stringent MAF (0.03) is used based on the PD population prevalence reported by de Lau and Breteler and Pringheim *et al* and the fact that incomplete penetrance is observed in PD cases (de Lau and Breteler, 2006; Pringsheim et al., 2014). Fifthly, given that these tools were created based on different algorithms which also accounts for the discrepancies observed among the tools in some cases, some of the true pathogenic variants, for which the tools did not agree, would be excluded from further analysis. Lastly, relatively few putative pathogenic variants were identified from our sequencing runs although our study participants exhibit familial PD, and this might be due to the small sample size. Furthermore, we are unable to fully estimate the contribution of the prioritized variants until we screen more PD patients and ethnically matched controls, especially for the variants that were detected in South African individuals belonging to unique ethnolinguistic groups.

Limitations for the *GBA* screening method are outlined in the manuscript. These include small sample size, absence of *GBA* enzyme activity tests to confirm the reduction of the beta-glucocerebrosidase enzyme to further support the possible pathogenicity of the variants, and lack of DNA of the affected family members to perform co-segregation analysis.

4.4 Future studies

Future studies that will derive from this work include using a CNV bioinformatics pipeline to identify structural rearrangements (Duan et al., 2013). Although CNV's are difficult to detect, they commonly occur in PD cases (La Cognata et al,. 2016). For instance, *SNCA* duplications and triplications, and *PRKN* exon rearrangements have been implicated in multiple families with familial PD. Another possibility is that Agilent's CNV detection panel OneSeq (De Witte et al., 2016) can be incorporated into the SureSelect custom gene panel if we are still unable to detect structural changes. Only two variants (in *PRKN* and *PINK1*) and the novel variants identified in the *GBA* manuscript section have been screened in ethnically matched controls (Bardien et al., 2009; Keyser et al., 2010a). Additionally, the *GBA* variants were validated with Sanger sequencing in the carriers. The pathogenicity of the remaining variants should also be assessed further by validating their presence with Sanger sequencing, screening in controls and performing protein modelling analyses. Alternatively, the variants may also be validated with Sanger sequencing as part of an on-going study of evaluating the quality or error rate of the sequencing runs on the Ion Torrent (Kubiritova et al., 2019). Additionally, newly discovered PD genes such as *PTRHD1,* and *PODXL* will be added to the gene panel (Kuipers et al., 2018; Sudhaman et al., 2016b). Recently, a *PTRHD1* 28 bp frameshift deletion has been shown to cause AR juvenile-onset parkinsonism and intellectual disability in a South African family of African ancestry (Kuipers et al., 2018). Furthermore, a frameshift variant in *PODXL* has been shown to cause AR juvenile-onset PD in two siblings of Indian ethnicity (Sudhaman et al., 2016b). Although the results have not been replicated in other populations it would be of benefit to screen for mutations in these genes especially in our AR juvenile-onset Atypical PD cases who do not have *PRKN, PINK1* or *DJ-1* mutations. Adding these genes will not significantly affect the cost of the gene panel, as we will remain in tier one pricing. Also, recruitment of the family members of individuals with the putative pathogenic variants should be done to assess if the variants co-segregate with the disease.

Although numerous tools for analysing NGS data have been developed, variant detection and prioritization for complex disorders is still a major challenge (Dashti and Gamieldien, 2017). Thus, we will also compare the reproducibility of a bioinformatic pipeline in our data set, specifically designed to detect variants in rare diseases using sequencing data from African populations (Schoonen et al., 2019). Although we validated the majority of the *GBA* variants identified with the gene panel using the specialized primers, we failed to detect the novel variant (p.S251L). Thus, we suspect that it is in the pseudogene. It is therefore imperative to assess the efficiency of the gene panel in detecting *GBA* variants versus using the Sanger sequencing method in the same set of samples to compare the reproducibility of both methods. Furthermore, all the *GBA* variants should also be assessed in larger numbers of PD cases and ethnically matched controls.

In addition, functional studies of the prioritized variants in *PINK1* and *PRKN* should be performed in PD cellular models to assess the consequences of these variants in pathways implicated in PD pathogenesis such

as mitochondrial function, apoptosis, lysosomal and protein clearing pathways. Functional studies of the *GBA* variants (p.L483P, p.R170L, p.D179H, and p.E365K) should also be performed, which includes measuring the beta glucocerebrosidase activity and stability in leukocytes, monocytes and patient-derived fibroblasts (Atashrazm et al., 2018; Ivanova et al., 2018; Nakagawa et al., 1982).

Lastly, although WES sequencing has many challenges its utility will be explored in our study populations especially in our unique South African ethnolinguistic groups whose pedigrees indicate a strong genetic component. Although families exhibiting a Mendelian inheritance pattern were chosen to establish if the known monogenic PD genes are contributors in our populations, few disease-causing variants in these genes were identified. This warrants the use of WES to identify candidate genes unique to our South African populations.

4.5 Conclusion

We successfully set up a relatively rapid technique for screening of the known PD genes as well as a method for screening of *GBA* in our local populations. With the gene panel, we were able to identify clinically relevant variants in the ARPD genes, *PRKN* and *PINK1*. Although the variants were heterozygous, we hypothesize that they may pose a risk for developing PD. Interestingly, we identified four pathogenic *GBA* variants of which they have been indicated to be pathogenic or to increase the risk of PD. Although *GBA* is known to be a major risk factor for developing PD, its contribution to South African populations has not been well established. To our knowledge, only one study has investigated *GBA* in PD patients of European ethnicity (Barkhuizen et al., 2017). Given the diversity of the South Africa population, more studies in all ethnicities are required to understand the genetic architecture of PD within South African PD patients. Studies such as this one are urgently needed to help guide genetic testing and also to identify cohorts of patients ready for gene-targeted precision clinical trials as a treatment for this devastating disorder.

References

Abou-Sleiman PM, Healy DG, Quinn N, Lees AJ, Wood NW. The role of pathogenic DJ-1 mutations in Parkinson's disease. Ann Neurol 2003;54:283–6. doi:10.1002/ana.10675.

Abou-Sleiman PM, Muqit MM, McDonald NQ, Yang YX, Gandhi S, Healy DG, et al. A heterozygous effect for PINK1 mutations in Parkinson's disease? Ann Neurol 2006;60:414–9. doi:10.1002/ana.20960.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9. doi:10.1038/nmeth0410-248.

Ahn BH, Rhim H, Kim SY, Sung YM, Lee MY, Choi JY, et al. alpha-Synuclein interacts with phospholipase D isozymes and inhibits pervanadate-induced phospholipase D activation in human embryonic kidney-293 cells. J Biol Chem 2002;277:12334–42. doi:10.1074/jbc.M110414200.

Alessi DR, Sammler E. LRRK2 kinase in Parkinson's disease. Science 2018;360:36–7. doi:10.1126/science.aar5683.

Alexander GE. Biology of Parkinson's disease: pathogenesis and pathophysiology of a multisystem neurodegenerative disorder. Dialogues Clin Neurosci 2004;6:259–80.

Alzohairy AM. Bioinformatics Training: Sequence Manipulation Suit SMS. Bioinforma Train 2014, p. 1–2.

Appel-Cresswell S, Rajput AH, Sossi V, Thompson C, Silva V, McKenzie J, et al. Clinical, positron emission tomography, and pathological studies of DNAJC13 p.N855S Parkinsonism. Mov Disord 2014;29:1684–7. doi:10.1002/mds.26019.

Araki M, Ito G, Tomita T. Physiological and pathological functions of LRRK2: implications from substrate proteins. Neuronal Signal 2018;2:NS20180005. doi:10.1042/NS20180005.

Arena G, Valente EM. PINK1 in the limelight: multiple functions of an eclectic protein in human health and disease. J Pathol 2017;241:251–63. doi:10.1002/path.4815.

Arndt S, Heitner R, Lane A, Ramsay M. Glucocerebrosidase gene mutations in black South Africans with Gaucher disease. Blood Cells Mol Dis 2009;43:129–33. doi:10.1016/J.BCMD.2009.02.008.

Atashrazm F, Hammond D, Perera G, Dobson-Stone C, Mueller N, Pickford R, et al. Reduced glucocerebrosidase activity in monocytes from patients with Parkinson's disease. Sci Rep 2018;8. doi:10.1038/s41598-018-33921-x.

Bajaj N, Hauser RA, Grachev ID. Clinical utility of dopamine transporter single-photon emission CT (DaT-SPECT) with (123I) ioflupane in diagnosis of parkinsonian syndromes. J Neurol Neurosurg Psychiatry 2013;84:1288–95. doi:10.1136/jnnp-2012-304436.

Baldereschi M, Di Carlo A, Rocca WA, Vanni P, Maggi S, Perissinotto E, et al. Parkinson's disease and parkinsonism in a longitudinal study: two-fold higher incidence in men. ILSA Working Group. Italian Longitudinal Study on Aging. Neurology 2000;55:1358–63. doi:10.1212/WNL.55.9.1358.

Balogou AA, Doh A, Grunitzky KE. Neurological disorders and endemic goiter: comparative analysis of 2 provinces in Togo. Bull Soc Pathol Exot 2001;94:406–10.

Balwani M, Grace ME, Desnick RJ. Gaucher disease: when molecular testing and clinical presentation disagree-the novel c.1226A>G(p.N370S)--RecNcil allele. J Inherit Metab Dis 2011;34:789–93. doi:10.1007/s10545-011-9307-7.

Bardien S., Keyser R, Lombard D, Du Plessis M, Human H, Carr J. Novel non-sense GCH1 mutation in a South African family diagnosed with dopa-responsive dystonia. Eur J Neurol 2010;17:510–2. doi:10.1111/j.1468-1331.2009.02725.x.

Bardien S, Keyser R, Yako Y, Lombard D, Carr J. Molecular analysis of the parkin gene in South African patients diagnosed with Parkinson's disease. Park Relat Disord 2009;15:116–21. doi:10.1016/j.parkreldis.2008.04.005.

Bardien Soraya, Marsberg A, Keyser R, Lombard D, Lesage S, Brice A, et al. LRRK2 G2019S mutation: Frequency and haplotype data in South African Parkinson's disease patients. J Neural Transm 2010;117:847–53. doi:10.1007/s00702-010-0423-6.

Barkhuizen M, Anderson DG, van der Westhuizen FH, Grobler AF. A molecular analysis of the GBA gene in Caucasian South Africans with Parkinson's disease. Mol Genet Genomic Med 2017;5:147–56. doi:10.1002/mgg3.267.

Baumann H, Wolff S, Münchau A, Hagenah JM, Lohmann K, Klein C. Evaluating the role of TMEM230 variants in Parkinson's disease. Park Relat Disord 2017;35:100–1. doi:10.1016/j.parkreldis.2016.12.015.

Behrens MI, Brüggemann N, Chana P, Venegas P, Kägi M, Parrao T, et al. Clinical spectrum of Kufor-Rakeb syndrome in the Chilean kindred with ATP13A2 mutations. Mov Disord 2010;25:1929–37. doi:10.1002/mds.22996.

Benabid AL. Deep brain stimulation for Parkinson's disease. Curr Opin Neurobiol 2003;13:696–706. doi:10.1016/J.CONB.2003.11.001.

Bentley DR. Whole-genome re-sequencing. Curr Opin Genet Dev 2006;16:545–52. doi:10.1016/J.GDE.2006.10.009.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456:53–9. doi:10.1038/nature07517.

Berardelli A, Rothwell JC, Thompson PD, Hallett M. Pathophysiology of bradykinesia in Parkinson's disease. Brain 2001;124:2131–46. doi:10.1093/brain/124.11.2131.

Beutler E. Gaucher disease as a paradigm of current issues regarding single gene mutations of humans. Proc Natl Acad Sci U S A 1993;90:5384–90. doi:10.1073/pnas.90.12.5384.

Beutler E, Gelbart T. Mutation analysis in gaucher disease. Am J Med Genet 1992;44:389–389. doi:10.1002/ajmg.1320440328.

Beutler E, Gelbart T, West C. Identification of Six New Gaucher Disease Mutations. Genomics 1993;15:203–5. doi:10.1006/geno.1993.1035.

Blocq P, Marinesco G. Sur un cas de tremblement parkinsonien hémiplégique symptomatique d'une tumeur du pédoncule cérébral. Comptes Rendus Séances Société Biol 1893;5:105–11.

Bogaerts V, Nuytemans K, Reumers J, Pals P, Engelborghs S, Pickut B, et al. Genetic variability in the mitochondrial serine protease HTRA2 contributes to risk for Parkinson disease. Hum Mutat 2008;29:832–40. doi:10.1002/humu.20713.

Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, et al. Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science 2003;299:256–9. doi:10.1126/science.1077209.

Braak H, Sandmann-Keil D, Gai W, Braak E. Extensive axonal Lewy neurites in Parkinson's disease: a novel pathological feature revealed by α-synuclein immunocytochemistry. Neurosci Lett 1999;265:67–9. doi:10.1016/S0304-3940(99)00208-6.

Braak H, Tredici K Del, Rüb U, de Vos RA., Jansen Steur EN., Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003;24:197–211. doi:10.1016/S0197-4580(02)00065-9.

Bridi JC, Hirth F. Mechanisms of α-Synuclein induced synaptopathy in parkinson's disease. Front Neurosci 2018;12. doi:10.3389/fnins.2018.00080.

Brissaud ÉA du texte. Leçons sur les maladies nerveuses. Salpêtrière. Paris, G. Masson; 1895.

Brookman-Amissah N. Increasing On-Target NGS Reads | GEN - Genetic Engineering and Biotechnology News. Mary Ann Liebert, Inc 2014. https://www.genengnews.com/magazine/220/increasing-on-target-ngs-reads/ (accessed October 21, 2019).

Burre J, Sharma M, Tsetsenis T, Buchman V, Etherton MR, Sudhof TC. Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. Science 2010;329:1663–7. doi:10.1126/science.1195227.

von Campenhausen S, Bornschein B, Wick R, Bötzel K, Sampaio C, Poewe W, et al. Prevalence and incidence of Parkinson's disease in Europe. Eur Neuropsychopharmacol 2005;15:473–90. doi:10.1016/J.euroneuro.2005.04.007.

Carlsson A. Treatment of Parkinson's with L-DOPA. The early discovery phase, and a comment on current problems. J Neural Transm 2002;109:777–87. doi:10.1007/s007020200064.

Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet 2017;49:1511–6. doi:10.1038/ng.3955.

Chang P, Fu Y, Zhao P, Wang C, Jiang M, Li R, et al. Variants in the 3' End of SLC6A3 in Northwest Han Population with Parkinson's. Parkinsons Dis 2019;2019:1–7. doi:10.1155/2019/6452471.

Charcot J. Lecons on, diseases of the nervous system. 1886.

Chartier-Harlin MC, Dachsel JC, Vilariño-Güell C, Lincoln SJ, Leprêtre F, Hulihan MM, et al. Translation Initiator EIF4G1 Mutations in Familial Parkinson Disease. Am J Hum Genet 2011;89:398–406. doi:10.1016/J.ajhg.2011.08.009.

Chen CM, Wu CH, Hsieh CH, Lin CH, Chen IC, Chen YC, et al. HTRA2 variations in Taiwanese Parkinson's disease. J Neural Transm 2014;121:491–8. doi:10.1007/s00702-013-1131-9.

Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 2015;31:2745–7. doi:10.1093/bioinformatics/btv195.

Choudhury A, Aron S, Sengupta D, Hazelhurst S, Ramsay M. African genetic diversity provides novel insights into evolutionary history and local adaptations. Hum Mol Genet 2018;27:R209–18. doi:10.1093/hmg/ddy161.

Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res 2009;19:1553–61. doi:10.1101/gr.092619.109.

Cilia R, Sironi F, Akpalu A, Cham M, Sarfo FS, Brambilla T, et al. Screening LRRK2 gene mutations in patients with Parkinson's disease in Ghana. J Neurol 2012;259:569–70. doi:10.1007/s00415-011-6210-y.

Clot F, Grabli D, Cazeneuve C, Roze E, Castelnau P, Chabrol B, et al. Exhaustive analysis of BH4 and dopamine biosynthesis genes in patients with Dopa-responsive dystonia. Brain 2009;132:1753–63. doi:10.1093/brain/awp084.

Conedera S, Apaydin H, Li Y, Yoshino H, Ikeda A, Matsushima T, et al. FBXO7 mutations in Parkinson's disease and multiple system atrophy. Neurobiol Aging 2016;40:192.e1-192.e5. doi:10.1016/J.neurobiolaging.2016.01.003.

Consortium The 1000 Genomes Project. A global reference for human genetic variation. Nature 2015;526:68–74. doi:10.1038/nature15393.

Corti O, Lesage S, Brice A. What genetics tells us about the causes and mechanisms of Parkinson's disease. Physiol Rev 2011;91:1161–218. doi:10.1152/physrev.00022.2010.

Crosiers D, Verstraeten A, Wauters E, Engelborghs S, Peeters K, Mattheijssens M, et al. Mutations in glucocerebrosidase are a major genetic risk factor for Parkinson's disease and increase susceptibility to dementia in a Flanders-Belgian cohort. Neurosci Lett 2016;629:160–4. doi:10.1016/j.neulet.2016.07.008.

Cui H, Kong Y, Zhang H. Oxidative stress, mitochondrial dysfunction, and aging. J Signal Transduct 2012;2012:646354. doi:10.1155/2012/646354.

Dagda RK, Chu CT. Mitochondrial quality control: insights on how Parkinson's disease related genes PINK1, parkin, and Omi/HtrA2 interact to maintain mitochondrial homeostasis. J Bioenerg Biomembr 2009;41:473–9. doi:10.1007/s10863-009-9255-1.

Dashti MJS, Gamieldien J. A practical guide to filtering and prioritizing genetic variants. Biotechniques 2017;62:18–30. doi:10.2144/000114492.

Davie CA. A review of Parkinson's disease. Br Med Bull 2008;86:109–27. doi:10.1093/bmb/ldn013.

Davydov EV., Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. PLoS Comput Biol 2010;6:e1001025. doi:10.1371/journal.pcbi.1001025.

Dawson TM, Dawson VL. The role of parkin in familial and sporadic Parkinson's disease. Mov Disord 2010;25. doi:10.1002/mds.22798.

Dawson TM, Dawson VL. Molecular pathways of neurodegeneration in Parkinson's disease. Science 2003;302:819–22. doi:10.1126/science.1087753.

Demirsoy S, Martin S, Motamedi S, van Veen S, Holemans T, Van den Haute C, et al. ATP13A2/PARK9 regulates endo-/lysosomal cargo sorting and proteostasis through a novel PI(3, 5)P2-mediated scaffolding function. Hum Mol Genet 2017;26:1656–69. doi:10.1093/hmg/ddx070.

Deng H, Gao K, Jankovic J. The VPS35 gene and Parkinson's disease. Mov Disord 2013;28:569–75. doi:10.1002/mds.25430.

Deng H, Le WD, Hunter CB, Ondo WG, Guo Y, Xie WJ, et al. Heterogeneous Phenotype in a Family With Compound Heterozygous Parkin Gene Mutations. Arch Neurol 2006;63:273. doi:10.1001/archneur.63.2.273.

Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. Ageing Res Rev 2018;42:72–85. doi:10.1016/J.arr.2017.12.007.

Deng H, Wu Y, Jankovic J. The EIF4G1 gene and Parkinson's disease. Acta Neurol Scand 2015a;132:73–8. doi:10.1111/ane.12397.

Deng H, Yuan L. Genetic variants and animal models in SNCA and Parkinson disease. Ageing Res Rev 2014;15:161–76. doi:10.1016/j.arr.2014.04.002.

Deng HX, Shi Y, Yang Y, Ahmeti KB, Miller N, Huang C, et al. Identification of TMEM230 mutations in familial Parkinson's disease. Nat Genet 2016;48:733–9. doi:10.1038/ng.3589.

Deuschl G, Wenzelburger R, Löffler K, Raethjen J, Stolze H. Essential tremor and cerebellar dysfunction Clinical and kinematic analysis of intention tremor. Brain 2000;123:1568–80. doi:10.1093/brain/123.8.1568.

van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet 2014;30:418–26. doi:10.1016/j.tig.2014.07.001.

Dilliott AA, Farhan SMK, Ghani M, Sato C, Liang E, Zhang M, et al. Targeted Next-generation Sequencing and Bioinformatics Pipeline to Evaluate Genetic Determinants of Constitutional Disease. J Vis Exp 2018. doi:10.3791/57266.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 2015;24:2125–37. doi:10.1093/hmg/ddu733.

Dotchin C, Msuya O, Kissima J, Massawe J, Mhina A, Moshy A, et al. The prevalence of Parkinson's disease in rural Tanzania. Mov Disord 2008;23:1567–672. doi:10.1002/mds.21898.

Drouet V, Lesage S. Synaptojanin 1 mutation in Parkinson's disease brings further insight into the neuropathological mechanisms. Biomed Res Int 2014;2014:289728. doi:10.1155/2014/289728.

Duan J, Zhang JG, Deng HW, Wang YP. Comparative Studies of Copy Number Variation Detection Methods for Next-Generation Sequencing Technologies. PLoS One 2013;8. doi:10.1371/journal.pone.0059128.

van Duijn CM, Dekker MCJ, Bonifati V, Galjaard RJ, Houwing-Duistermaat JJ, Snijders PJ, et al. PARK7, a Novel Locus for Autosomal Recessive Early-Onset Parkinsonism, on Chromosome 1p36. Am J Hum Genet 2001;69:629–34. doi:10.1086/322996.

Duran R, Mencacci NE, Angeli AV., Shoai M, Deas E, Houlden H, et al. The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. Mov Disord 2013;28:232–6. doi:10.1002/mds.25248.

Dvir H, Harel M, McCarthy AA, Toker L, Silman I, Futerman AH, et al. X-ray structure of human acid-β-glucosidase, the defective enzyme in Gaucher disease. EMBO Rep 2003;4:704–9. doi:10.1038/sj.embor.embor873.

Dyment DA, Smith AC, Humphreys P, Schwartzentruber J, Beaulieu CL, Bulman DE, et al. Homozygous nonsense mutation in SYNJ1 associated with intractable epilepsy and tau pathology. Neurobiol Aging 2015;36:1222.e1-1222.e5. doi:10.1016/J.NEUROBIOLAGING.2014.09.005.

Edvardson S, Cinnamon Y, Ta-Shma A, Shaag A, Yim YI, Zenvirt S, et al. A deleterious mutation in DNAJC6 encoding the neuronal-specific clathrin-uncoating Co-chaperone auxilin, is associated with juvenile parkinsonism. PLoS One 2012;7:e36458. doi:10.1371/journal.pone.0036458.

El-Metwally S, Hamza T, Zakaria M, Helmy M. Next-generation sequence assembly: four stages of data processing and computational challenges. PLoS Comput Biol 2013;9:e1003345. doi:10.1371/journal.pcbi.1003345.

Elbaz A, Nichols E, Abd-Allah F, Abdelalim A, Adsuar JC, Ansha MG, et al. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol 2018;17:939–53. doi:10.1016/S1474-4422(18)30295-3.

Ellingford JM, Campbell C, Barton S, Bhaskar S, Gupta S, Taylor RL, et al. Validation of copy number variation analysis for next-generation sequencing diagnostics. Eur J Hum Genet 2017;25:719–24. doi:10.1038/ejhg.2017.42.

Emamzadeh FN. Alpha-synuclein structure, functions, and interactions. J Res Med Sci 2016;21:29. doi:10.4103/1735-1995.181989.

Engelender S, Isacson O. The Threshold Theory for Parkinson's Disease. Trends Neurosci 2017;40:4–14. doi:10.1016/J.TINS.2016.10.008.

Eyal N, Firon N, Wilder S, Kolodny EH, Horowitz M. Three unique base pair changes in a family with Gaucher disease. Hum Genet 1991;87:328–32. doi:10.1007/BF00200914.

Factor SA, Weiner WJ. Prior history of head trauma in Parkinson's disease. Mov Disord 1991;6:225–9. doi:10.1002/mds.870060306.

Fahn S, Elton R. UPDRS Development Committee. The Unified Parkinson's Disease Rating Scale. Macmillan Healthc Inf 1987:153–63.

Fearnley JM, Lees AJ. Ageing and Parkinson's disease: substantia nigra regional selectivity. Brain 1991;114:2283–301. doi:10.1093/brain/114.5.2283.

Fedarko NS. The biology of aging and frailty. Clin Geriatr Med 2011;27:27–37. doi:10.1016/j.cger.2010.08.006.

Ferese R, Modugno N, Campopiano R, Santilli M, Zampatti S, Giardina E, et al. Four Copies of SNCA Responsible for Autosomal Dominant Parkinson's Disease in Two Italian Siblings. Parkinsons Dis 2015;2015:1–6. doi:10.1155/2015/546462.

Ferreira M, Massano J. An updated review of Parkinson's disease genetics and clinicopathological correlations. Acta Neurol Scand 2017;135:273–84. doi:10.1111/ane.12616.

Di Fonzo A, Chien HF, Socal M, Giraudo S, Tassorelli C, Iliceto G, et al. ATP13A2 missense mutations in juvenile parkinsonism and young onset Parkinson disease. Neurology 2007;68:1557–62. doi:10.1212/01.wnl.0000260963.08711.08.

Di Fonzo A, Dekker MCJ, Montagna P, Baruzzi A, Yonova EH, Correia Guedes L, et al. FBXO7 mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. Neurology 2009;72:240–5. doi:10.1212/01.wnl.0000338144.10967.2b.

Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nat Rev Genet 2009;10:241–51. doi:10.1038/nrg2554.

Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol 2002;51:296–301. doi:10.1002/ana.10113.

Funayama M, Ohe K, Amo T, Furuya N, Yamaguchi J, Saiki S, et al. CHCHD2 mutations in autosomal dominant late-onset Parkinson's disease: a genome-wide linkage and sequencing study. Lancet Neurol 2015;14:274–82. doi:10.1016/S1474-4422(14)70266-2.

Gagliardi M, Annesi G, Procopio R, Morelli M, Iannello G, Bonapace G, et al. DNAJC13 mutation screening in patients with Parkinson's disease from South Italy. Parkinsonism Relat Disord 2018;55:134–7. doi:10.1016/j.parkreldis.2018.06.004.

Gan-Or Z, Amshalom I, Kilarski LL, Bar-Shira A, Gana-Weisz M, Mirelman A, et al. Differential effects of severe vs mild GBA mutations on Parkinson disease. Neurology 2015;84:880. doi:10.1212/WNL.0000000000001315.

Gandhi S, Wood NW. Molecular pathogenesis of Parkinson's disease. Hum Mol Genet 2005;14:2749–55. doi.org/10.1093/hmg/ddi308.

Gatto EM, Aldinio V. Impulse Control Disorders in Parkinson's Disease. A Brief and Comprehensive Review. Front Neurol 2019;10:351. doi:10.3389/fneur.2019.00351.

Gaucher P. Primary epithelioma of the spleen, idiopathic hypertrophy of the spleen without leukemia. Paris Fac Med Paris 1882.

Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry 1988;51:745–52. doi:10.1136/jnnp.51.6.745.

Gibb WR, Narabayashi H, Yokochi M, Iizuka R, Lees AJ. New pathologic observations in juvenile onset parkinsonism with dystonia. Neurology 1991;41:820–2. doi:10.1212/wnl.41.6.820.

Gibb WR, Lees AJ. A comparison of clinical and pathological features of young- and old-onset parkinson's disease. Neurology 1988;38:1402–6. doi:10.1212/wnl.38.9.1402.

Giri A, Mok KY, Jansen I, Sharma M, Tesson C, Mangone G, et al. Lack of evidence for a role of genetic variation in TMEM230 in the risk for Parkinson's disease in the Caucasian population. Neurobiol Aging 2017;50:167.e11-167.e13. doi:10.1016/j.neurobiolaging.2016.10.004.

Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations. Mov Disord 2004;19:1020–8. doi:10.1002/mds.20213.

Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord 2008;23:2129–70. doi:10.1002/mds.22340.

Goker-Alpan O, Schiffmann R, Park JK, Stubblefield BK, Tayebi N, Sidransky E. Phenotypic continuum in neuronopathic gaucher disease: an intermediate phenotype between type 2 and type 3. J Pediatr 2003;143:273–6. doi:10.1067/S0022-3476(03)00302-0.

Gorostidi A, Martí-Massó JF, Bergareche A, Rodríguez-Oroz MC, López de Munain A, Ruiz-Martínez J. Genetic Mutation Analysis of Parkinson's Disease Patients Using Multigene Next-Generation Sequencing Panels. Mol Diagnosis Ther 2016;20:481–91. doi:10.1007/s40291-016-0216-1.

Grace ME, Ashton-Prolla P, Pastores GM, Soni A, Desnick RJ. Non-pseudogene-derived complex acid β-glucosidase mutations causing mild type 1 and severe type 2 Gaucher disease. J Clin Invest 1999;103:817–23. doi:10.1172/JCI5168.

Gray CW, Ward RV, Karran E, Turconi S, Rowles A, Viglienghi D, et al. Characterization of human HtrA2, a novel serine protease involved in the mammalian cellular stress response. Eur J Biochem 2000;267:5699–710. doi:10.1172/JCI5168.

Gregory A, Polster BJ, Hayflick SJ. Clinical and genetic delineation of neurodegeneration with brain iron accumulation. J Med Genet 2009;46:73–80. doi:10.1136/jmg.2008.061929.

Guella I, Sherman HE, Appel-Cresswell S, Rajput A, Rajput AH, Farrer MJ. Parkinsonism in GTP cyclohydrolase 1 mutation carriers. Brain 2015;138:e349. doi:10.1093/brain/awu341.

Gustavsson EK, Trinh J, Guella I, Vilariño-Güell C, Appel-Cresswell S, Stoessl AJ, et al. DNAJC13 genetic variants in parkinsonism. Mov Disord 2015;30:273–8. doi:10.1002/mds.26064.

Haehner A, Hummel T, Hummel C, Sommer U, Junghanns S, Reichmann H. Olfactory loss may be a first sign of idiopathic Parkinson's disease. Mov Disord 2007;22:839–42. doi:10.1002/mds.21413.

Hall TA. A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. Nucleic Acids Symp 1999;41:95–8.

Hasin-Brumshtein Y, Celeste M, Ramirez M, Arbiza L, Zeitoun R. The Importance of Coverage Uniformity Over On-Target Rate for Efficient Targeted NGS. 2018.

Healy D, Falchi M, O'Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. Lancet Neurol 2008;7:583–90. doi:10.1016/S1474-4422(08)70117-0.

Hindle JV. Ageing, neurodegeneration and Parkinson's disease. Age Ageing 2010;39:156–61. doi:10.1093/ageing/afp223.

Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. Neurology 1967;17:427–42. doi:10.1212/wnl.17.5.427.

Holleran WM, Ginns EI, Menon GK, Grundmann JU, Fartasch M, McKinney CE, et al. Consequences of beta-glucocerebrosidase deficiency in epidermis. Ultrastructure and permeability barrier alterations in Gaucher disease. J Clin Invest 1994;93:1756–64. doi:10.1172/JCI117160.

Hornykiewicz O. Basic Research on Dopamine in Parkinson's Disease and the Discovery of the Nigrostriatal Dopamine Pathway: The View of an Eyewitness. Neurodegener Dis 2008;5:114–7. doi:10.1159/000113678.

Horowitz M, Pasmanik-Chor M, Ron I, Kolodny EH. The enigma of the E326K mutation in acid β-glucocerebrosidase. Mol Genet Metab 2011;104:35–8. doi:10.1016/j.ymgme.2011.07.002.

Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E. The human glucocerebrosidase gene and pseudogene: Structure and evolution. Genomics 1989;4:87–96. doi:10.1016/0888-7543(89)90319-4.

Hruska KS, LaMarca ME, Scott CR, Sidransky E. Gaucher disease: mutation and polymorphism spectrum in the glucocerebrosidase gene (GBA). Hum Mutat 2008;29:567–83. doi:10.1002/humu.20676.

Hu J, Ng PC. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. PLoS One 2013;8:e77940. doi:10.1371/journal.pone.0077940.

Huang Y, Deng L, Zhong Y, Yi M. The association between E326K of GBA and the risk of Parkinson's disease. Parkinsons Dis 2018;2018. doi:10.1155/2018/1048084.

Hughes AJ, Daniel SE, Blankson S, Lees AJ. A Clinicopathologic Study of 100 Cases of Parkinson's Disease. Arch Neurol 1993;50:140–8. doi:10.1001/archneur.1993.00540020018011.

Hughes AJ, Daniel SE, Lees AJ. Neurology. Neurology 2001;42:1142–1142. doi:10.1212/wnl.57.8.1497.

Iannello G, Graziano C, Cenacchi G, Cordelli DM, Zuntini R, Papa V, et al. A new PLA2G6 mutation in a family with infantile neuroaxonal dystrophy. J Neurol Sci 2017;381:209–12. doi:10.1016/j.jns.2017.08.3260.

Ikeda A, Matsushima T, Daida K, Nakajima S, Conedera S, Li Y, et al. A novel mutation of CHCHD2 p.R8H in a sporadic case of Parkinson's disease. Park Relat Disord 2017;34:66–8. doi:10.1016/j.parkreldis.2016.10.018.

Inestrosa NC, Marzolo MP, Bonnefont AB. Cellular and molecular basis of estrogen's neuroprotection. Mol Neurobiol 1998;17:73–86. doi:10.1007/BF02802025.

Inoshita T, Arano T, Hosaka Y, Meng H, Umezaki Y, Kosugi S, et al. Vps35 in cooperation with LRRK2 regulates synaptic vesicle endocytosis through the endosomal pathway in Drosophila. Hum Mol Genet 2017;26:2933–48. doi:10.1093/hmg/ddx179.

Ivanova MM, Changsila E, Turgut A, Goker-Alpan O. Individualized screening for chaperone activity in gaucher disease using multiple patient derived primary cell lines. Am J Transl Res 2018;10:3750–61. doi:10.1016/j.ymgme.2017.12.170.

Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, et al. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 2016;48:1581–6. doi:10.1038/ng.3703.

Jankovic J. Parkinson's disease: clinical features and diagnosis. J Neurol Neurosurg Psychiatry 2008;79:368–76. doi:10.1136/jnnp.2007.131045.

Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagnostics 2017;19:341–65. doi:10.1016/j.jmoldx.2017.01.011.

Jesús S, Gómez-Garre P, Carrillo F, Cáceres-Redondo MT, Huertas-Fernández I, Bernal-Bernal I, et al. Analysis of c.801-2A>G mutation in the DNAJC6 gene in Parkinson's disease in southern Spain. Parkinsonism Relat Disord 2014;20:248–9. doi:10.1016/j.parkreldis.2013.10.018.

John JS, Quinn TW. Rapid capture of DNA targets. Biotechniques 2008;44:259–64. doi:10.2144/000112633.

Joseph S, Schulz JB, Stegmüller J. Mechanistic contributions of FBXO7 to Parkinson disease. J Neurochem 2018;144:118–27. doi:10.1111/jnc.14253.

Kapa biosystems. High-Throughput NGS Library Preparation Technical Guide - Google Search. 2014. https://www.kapabiosystems.com/document/high-throughput-ngs-library-preparation-technical-guide/ (accessed October 21, 2019).

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. BioRxiv 2019:531210. doi:10.1101/531210.

Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res 2017;45:D840–5. doi:10.1093/nar/gkw971.

Karkheiran S, Shahidi GA, Walker RH, Paisán-Ruiz C. PLA2G6-associated Dystonia-Parkinsonism: Case Report and Literature Review. Tremor Other Hyperkinet Mov (NY) 2015;5:317. doi:10.7916/D84Q7T4W.

Kasten M, Hartmann C, Hampf J, Schaake S, Westenberger A, Vollstedt EJ, et al. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin , PINK1 , DJ1: MDSGene Systematic Review. Mov Disord 2018;33:730–41. doi:10.1002/mds.27352.

Kasten M, Klein C. The many faces of alpha-synuclein mutations. Mov Disord 2013;28:697–701. doi:10.1002/mds.25499.

Kazlauskaite A, Muqit MM. PINK1 and Parkin – mitochondrial interplay between phosphorylation and ubiquitylation in Parkinson's disease. FEBS J 2015;282:215–23. doi:10.1111/febs.13127.

Keus SH, Bloem BR, Hendriks EJ, Bredero-Cohen AB, Munneke M. Evidence-based analysis of physical therapy in Parkinson's disease with recommendations for practice and research. Mov Disord 2007;22:451–60. doi:10.1002/mds.21244.

Keyser RJ, Lesage S, Brice A, Carr J, Bardien S. Assessing the prevalence of PINK1 genetic variants in South African patients diagnosed with early- and late-onset Parkinson's disease. Biochem Biophys Res Commun 2010a;398:125–9. doi:10.1016/j.bbrc.2010.06.049.

Keyser RJ, Lombard D, Veikondis R, Carr J, Bardien S. Analysis of exon dosage using MLPA in South African Parkinson's disease patients. Neurogenetics 2010b;11:305–12. doi:10.1007/s10048-009-0229-6.

Kilarski LL, Pearson JP, Newsway V, Majounie E, Knipe MD, Misbahuddin A, et al. Systematic Review and UK-Based Study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov Disord 2012;27:1522–9. doi:10.1002/mds.25132.

Kinghorn KJ, Castillo-Quan JI, Bartolome F, Angelova PR, Li L, Pope S, et al. Loss of PLA2G6 leads to elevated mitochondrial lipid peroxidation and mitochondrial dysfunction. Brain 2015;138:1801–16. doi:10.1093/brain/awv132.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 2014;46:310–5. doi:10.1038/ng.2892.

Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 1998;392:605–8. doi:10.1038/33416.

Klein C, Lohmann-Hedrich K, Rogaeva E, Schlossmacher MG, Lang AE. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. Lancet Neurol 2007;6:652–62. doi:10.1016/S1474-4422(07)70174-6.

Kong N, Thao K, Huang C, Weimer BC, Appel M, Lappin S, et al. Automated Library Construction Using KAPA Library Preparation Kits on the Agilent NGS Workstation Yields High-Quality Libraries for Whole-Genome Sequencing on the Illumina Platform. Agil Technol Appl Note 2015. doi:doi:10.6084/m9.figshare.1386854.

Kono S, Shirakawa K, Ouchi Y, Sakamoto M, Ida H, Sugiura T, et al. Dopaminergic neuronal dysfunction associated with parkinsonism in both a Gaucher disease patient and a carrier. J Neurol Sci 2007;252:181–4. doi:10.1016/j.jns.2006.10.019.

Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. Bioinformatics 2007;23:1289–91. doi:10.1093/bioinformatics/btm091.

Köroğlu Ç, Baysal L, Cetinkaya M, Karasoy H, Tolun A. DNAJC6 is responsible for juvenile parkinsonism with phenotypic variability. Parkinsonism Relat Disord 2013;19:320–4. doi:10.1016/j.parkreldis.2012.11.006.

Kozarewa I, Armisen J, Gardner AF, Slatko BE, Hendrickson CL. Overview of Target Enrichment Strategies. Curr Protoc Mol Biol 2015;112:7.1-7. doi:10.1002/0471142727.mb0721s112.

Krebs CE, Karkheiran S, Powell JC, Cao M, Makarov V, Darvish H, et al. The Sac1 domain of SYNJ1 identified mutated in a family with early-onset progressive Parkinsonism with generalized seizures. Hum Mutat 2013;34:1200–7. doi:10.1002/humu.22372.

Krüger R, Kuhn W, Leenders KL, Sprengelmeyer R, Müller T, Woitalla D, et al. Familial parkinsonism with synuclein pathology: clinical and PET studies of A30P mutation carriers. Neurology 2001;56:1355–62. doi:10.1212/wnl.56.10.1355.

Kubiritova Z, Gyuraszova M, Nagyova E, Hyblova M, Harsanyova M, Budis J, et al. On the critical evaluation and confirmation of germline sequence variants identified using massively parallel sequencing. J Biotechnol 2019;298:64–75. doi:10.1016/j.jbiotec.2019.04.013.

Kuipers DJS, Carr J, Bardien S, Thomas P, Sebate B, Breedveld GJ, et al. PTRHD1 Loss-of-function mutation in an african family with juvenile-onset Parkinsonism and intellectual disability. Mov Disord 2018;33:1814–9. doi:10.1002/mds.27501.

Kumar A, Tamjar J, Waddell AD, Woodroof HI, Raimi OG, Shaw AM, et al. Structure of PINK1 and mechanisms of Parkinson's disease-associated mutations. Elife 2017;6. doi:10.7554/eLife.29985.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009;4:1073–81. doi:10.1038/nprot.2009.86.

Kurian MA, Morgan NV, MacPherson L, Foster K, Peake D, Gupta R, et al. Phenotypic spectrum of neurodegeneration associated with mutations in the PLA2G6 gene (PLAN). Neurology 2008;70:1623–9. doi:10.1212/01.wnl.0000310986.48286.8e.

Kurian MA, Zhen J, Cheng SY, Li Y, Mordekar SR, Jardine P, et al. Homozygous loss-of-function mutations in the gene encoding the dopamine transporter are associated with infantile parkinsonism-dystonia. J Clin Invest 2009;119:1595–603. doi:10.1172/JCI39060.

La Cognata V, D'Agata V, Cavalcanti F, Cavallaro S. Genetics of Parkinson's Disease: The Role of Copy Number Variations, Challenges in Parkinson's Disease. IntechOpen 2016; doi: 10.5772/62881.

Lampariello LR, Cortelazzo A, Guerranti R, Sticozzi C, Valacchi G. The Magic Velvet Bean of Mucuna pruriens. J Tradit Complement Med 2012;2:331–9. doi:10.1016/s2225-4110(16)30119-5.

Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46:D1062–7. doi:10.1093/nar/gkx1153.

de Lau LM, Breteler MM. Epidemiology of Parkinson's disease. Lancet Neurol 2006;5:525–35. doi:10.1016/S1474-4422(06)70471-9.

Lautier C, Goldwurm S, Dürr A, Giovannone B, Tsiaras WG, Pezzoli G, et al. Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease. Am J Hum Genet 2008;82:822–33. doi:10.1016/j.ajhg.2008.01.015.

Lekoubou A, Echouffo-Tcheugui JB, Kengne AP. Epidemiology of neurodegenerative diseases in sub-Saharan Africa: a systematic review. BMC Public Health 2014;14:653. doi:10.1186/1471-2458-14-653.

Lesage S, Anheim M, Condroyer C, Pollak P, Durif F, Dupuits C, et al. Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. Hum Mol Genet 2011;20:202–10. doi:10.1093/hmg/ddq454.

Lesage S, Condroyer C, Hecham N, Anheim M, Belarbi S, Lohman E, et al. Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa. Neurology 2011;76:301–3. doi:10.1212/WNL.0b013e318207b01e.

Lesage S, Condroyer C, Lannuzel A, Lohmann E, Troiano A, Tison F, et al. Molecular analyses of the LRRK2 gene in European and North African autosomal dominant Parkinson's disease. J Med Genet 2009;46:458–64. doi:10.1136/jmg.2008.062612.

Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. Am J Hum Genet 2016;98:500–13. doi:10.1016/j.ajhg.2016.01.014.

Lesage S, Dürr A, Tazir M, Lohmann E, Leutenegger AL, Janin S, et al. LRRK2 G2019S as a cause of Parkinson's Disease in North African Arabs. N Engl J Med 2006;354:422–3. doi:10.1056/NEJMc055540.

Lewy F. Paralysis agitans. I. Pathologische Anatomie. Handb Der Neurol 1912;3:920–58.

Li H. Exploring single-sample SNP and INDEL calling with whole-genome de novo assembly. Bioinformatics 2012;28:1838–44. doi:10.1093/bioinformatics/bts280.

Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 2010;26:589–95. doi:10.1093/bioinformatics/btp698.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. doi:10.1093/bioinformatics/btp324.

Li NN, Wang L, Tan EK, Cheng L, Sun XY, Lu ZJ, et al. Genetic analysis of CHCHD2 gene in Chinese Parkinson's disease. Am J Med Genet Part B Neuropsychiatr Genet 2016;171:1148–52. doi:10.1002/ajmg.b.32498.

Lin CH, Chen ML, Chen GS, Tai CH, Wu RM. Novel variant Pro143Ala in HTRA2 contributes to Parkinson's disease by inducing hyperphosphorylation of HTRA2 protein in mitochondria. Hum Genet 2011;130:817–27. doi:10.1007/s00439-011-1041-6.

Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat 2016;37:235–41. doi:10.1002/humu.22932.

Loane C, Politis M. Positron emission tomography neuroimaging in Parkinson's disease. Am J Transl Res 2011;3:323–41.

Lohmann E, Coquel AS, Honoré A, Gurvit H, Hanagasi H, Emre M, et al. A new F-box protein 7 gene mutation causing typical Parkinson's disease. Mov Disord 2015;30:1130–3. doi:10.1002/mds.26266.

Loman NJ, Misra RV., Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 2012;30:434–9. doi:10.1038/nbt.2198.

Lücking CB, Dürr A, Bonifati V, Vaughan J, De Michele G, Gasser T, et al. Association between Early-Onset Parkinson's Disease and Mutations in the Parkin Gene. N Engl J Med 2000;342:1560–7. doi:10.1056/NEJM200005253422103.

Lunati A, Lesage S, Brice A. The genetic landscape of Parkinson's disease. Rev Neurol (Paris) 2018;174:628–43. doi:10.1016/j.neurol.2018.08.004.

Ma D, Foo JN, Yulin Ng E, Zhao Y, Liu JJ, Tan EK. Screening for TMEM230 mutations in young-onset Parkinson's disease. Neurobiol Aging 2017;58:239.e9-239.e10. doi:10.1016/j.neurobiolaging.2017.06.011.

Malek N, Weil RS, Bresner C, Lawton MA, Grosset KA, Tan M, et al. Features of GBA-associated Parkinson's disease at presentation in the UK Tracking Parkinson's study. J Neurol Neurosurg Psychiatry 2018;89:702–9. doi:10.1136/jnnp-2017-317348.

Malini E, Grossi S, Deganuto M, Rosano C, Parini R, Dominisini S, et al. Functional analysis of 11 novel GBA alleles. Eur J Hum Genet 2014;22:511–6. doi:10.1038/ejhg.2013.182.

Mandel H, Saita S, Edvardson S, Jalas C, Shaag A, Goldsher D, et al. Deficiency of HTRA2/Omi is associated with infantile neurodegeneration and 3-methylglutaconic aciduria. J Med Genet 2016;53:690–6. doi:10.1136/jmedgenet-2016-103922.

Manyam BV. Paralysis agitans and levodopa in "Ayurveda": ancient Indian medical treatise. Mov Disord 1990;5:47–8. doi:10.1002/mds.870050112.

Mardis ER. A decade's perspective on DNA sequencing technology. Nature 2011;470:198–203. doi:10.1038/nature09796.

Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 2008;9:387–402. doi:10.1146/annurev.genom.9.081307.164359.

Martin I, Kim JW, Dawson VL, Dawson TM. LRRK2 pathobiology in Parkinson's disease. J Neurochem 2014;131:554–65. doi:10.1111/jnc.12949.

Martins LM, Morrison A, Klupsch K, Fedele V, Moisoi N, Teismann P, et al. Neuroprotective role of the Reaper-related serine protease HtrA2/Omi revealed by targeted deletion in mice. Mol Cell Biol 2004;24:9848–62. doi:10.1128/MCB.24.22.9848-9862.2004.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol 2016;17:122. doi:10.1186/s13059-016-0974-4.

McNeill A, Wu RM, Tzen KY, Aguiar PC, Arbelo JM, Barone P, et al. Dopaminergic Neuronal Imaging in Genetic Parkinson's Disease: Insights into Pathogenesis. PLoS One 2013;8. doi:10.1371/journal.pone.0069190.

Meade RM, Fairlie DP, Mason JM. Alpha-synuclein structure and Parkinson's disease – lessons and emerging principles. Mol Neurodegener 2019;14. doi:10.1186/s13024-019-0329-1.

Meeus B, Verstraeten A, Crosiers D, Engelborghs S, Van den Broeck M, Mattheijssens M, et al. DLB and PDD: A role for mutations in dementia and Parkinson disease genes? Neurobiol Aging 2012;33:629.e5-629.e18. doi:10.1016/j.neurobiolaging.2011.10.014.

Menashe-Oren A, Stecklov G. "IFAD RESEARCH SERIES 17 - Population age structure and sex composition in sub-Saharan Africa. IFAD Research Series 2017;280055. doi:10.22004/ag.econ.280055.

Mencacci NE, Isaias IU, Reich MM, Ganos C, Plagnol V, Polke JM, et al. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. Brain 2014;137:2480–92. doi:10.1093/brain/awu179.

van der Merwe C, Jalali Sefid Dashti Z, Christoffels A, Loos B, Bardien S. Evidence for a common biological pathway linking three Parkinson's disease-causing genes: parkin , PINK1 and DJ-1. Eur J Neurosci 2015;41:1113–25. doi:10.1111/ejn.12872.

Mohan M, Mellick GD. Role of the VPS35 D620N mutation in Parkinson's disease. Parkinsonism Relat Disord 2017;36:10–8. doi:10.1016/J.parkreldis.2016.12.001.

Mohan V, Radha V. Precision Diabetes Is Slowly Becoming a Reality. Med Princ Pract 2019;28:1–9. doi:10.1159/000497241.

Montfort M, Chabás A, Vilageliu L, Grinberg D. Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: Pathogenic changes and "modifier" polymorphisms. Hum Mutat 2004;23:567–75. doi:10.1002/humu.20043.

Morgan NV., Westaway SK, Morton JE, Gregory A, Gissen P, Sonek S, et al. PLA2G6, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. Nat Genet 2006;38:752–4. doi:10.1038/ng1826.

Nakagawa S, Kumin S, Sachs G, Nitowsky HM. Heterozygote detection of Type I Gaucher disease using blood platelets. Clin Chim Acta 1982;118:99–107. doi:10.1016/0009-8981(82)90231-5.

Nalls MA, Duran R, Lopez G, Kurzawa-Akanbi M, McKeith IG, Chinnery PF, et al. A multicenter study of glucocerebrosidase mutations in dementia with Lewy bodies. JAMA Neurol 2013;70:727–35. doi:10.1001/jamaneurol.2013.1925.

Nandipati S, Litvan I. Environmental Exposures and Parkinson's Disease. Int J Environ Res Public Health 2016;13. doi:10.3390/ijerph13090881.

Ning Z, Cox AJ, Mullikin JC. SSAHA: A Fast Search Method for Large DNA Databases. Genome Res 2001;11:1725–9. doi:10.1101/gr.194201.

Nishioka K, Vilariño-Güell C, Cobb SA, Kachergus JM, Ross OA, Wider C, et al. Glucocerebrosidase mutations are not a common risk factor for Parkinson disease in North Africa. Neurosci Lett 2010;477:57. doi:10.1016/j.neulet.2009.11.066.

Oczkowska A, Florczak-Wyspianska J, Permoda-Osip A, Owecki M, Lianeri M, Kozubski W, et al. Analysis of PRKN Variants and Clinical Features in Polish Patients with Parkinson's Disease. Curr Genomics 2015;16:215–23. doi:10.2174/1389202916666150326002549.

Okubadejo N, Britton A, Crews C, Akinyemi R, Hardy J, Singleton A, et al. Analysis of Nigerians with apparently sporadic Parkinson disease for mutations in LRRK2, PRKN and ATXN3. PLoS One 2008;3. doi:10.1371/journal.pone.0003421.

Okun MS. Deep-Brain Stimulation for Parkinson's Disease. N Engl J Med 2012;367:1529–38. doi:10.1056/NEJMct1208070.

Olanow CW, Agid Y, Mizuno Y, Albanese A, Bonucelli U, Damier P, et al. Levodopa in the treatment of Parkinson's disease: Current controversies. Mov Disord 2004;19:997–1005. doi:10.1002/mds.20243.

Olgiati S, Quadri M, Fang M, Rood JP, Saute JA, Chien HF, et al. DNAJC6 Mutations Associated With Early-Onset Parkinson's Disease. Ann Neurol 2016;79:244–56. doi:10.1002/ana.24553.

Olszewska DA, Fearon C, Lynch T. Novel gene (TMEM230) linked to Parkinson's disease. J Clin Mov Disord 2016;3:17. doi:10.1186/s40734-016-0046-7.

Oluwole OG. Implementation of Targeted Resequencing Strategies to Identify Pathogenic Mutations in Nigerian and South African Patients with Parkinson ' s Disease. 2019.

Osuntokun B, Adeuja A, Schoenberg B, Bademosi O, Nottidge V, Olumide A, et al. Neurological disorders in Nigerian Africans: a community-based study. Acta Neurol Scand 1987;75:13–21. doi:10.1111/j.1600-0404.1987.tb07883.x.

Ovallath S, Deepa P. The history of parkinsonism: Descriptions in ancient Indian medical literature. Mov Disord 2013;28:566–8. doi:10.1002/mds.25420.

Paisan-Ruiz C, Bhatia KP, Li A, Hernandez D, Davis M, Wood NW, et al. Characterization of PLA2G6 as a locus for dystonia-parkinsonism. Ann Neurol 2009;65:19–23. doi:10.1002/ana.21415.

Paisán-Ruiz C, Guevara R, Federoff M, Hanagasi H, Sina F, Elahi E, et al. Early-onset L-dopa-responsive parkinsonism with pyramidal signs due to ATP13A2, PLA2G6, FBXO7 and spatacsin mutations. Mov Disord 2010;25:1791–800. doi:10.1002/mds.23221.

Paisán-Ruíz C, Jain S, Evans EW, Gilks WP, Simón J, van der Brug M, et al. Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease. Neuron 2004;44:595–600. doi:10.1016/J.neuron.2004.10.023.

Pan T, Kondo S, Le W, Jankovic J. The role of autophagy-lysosome pathway in neurodegeneration associated with Parkinson's disease. Brain 2008;131:1969–78. doi:10.1093/brain/awm318.

Pandey S, Srivanitchapoom P. Levodopa-induced dyskinesia: Clinical features, pathophysiology, and medical management. Ann Indian Acad Neurol 2017;20:190–8. doi:10.4103/aian.AIAN_239_17.

Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, et al. Significant linkage of Parkinson disease to chromosome 2q36-37. Am J Hum Genet 2003;72:1053–7. doi:10.1086/374383.

Pankratz N, Nichols WC, Uniacke SK, Halter C, Rudolph A, Shults C, et al. Genome screen to identify susceptibility genes for Parkinson disease in a sample without parkin mutations. Am J Hum Genet 2002;71:124–35. doi:10.1086/341282.

Parkinson J. An essay on the shaking palsy. J Neuropsychiatry Clin Neurosci 1817;14. doi:10.1176/jnp.14.2.223.

Parkkinen L, Pirttilä T, Alafuzoff I. Applicability of current staging/categorization of α-synuclein pathology and their clinical relevance. Acta Neuropathol 2008;115:399–407. doi:10.1007/s00401-008-0346-6.

Pasanen P, Myllykangas L, Siitonen M, Raunio A, Kaakkola S, Lyytinen J, et al. A novel α-synuclein mutation A53E associated with atypical multiple system atrophy and Parkinson's disease-type pathology. Neurobiol Aging 2014;35:2180.e1-2180.e5. doi:10.1016/j.neurobiolaging.2014.03.024.

Pastores GM, Hughes DA. Gaucher Disease. University of Washington, Seattle; 1993. doi:https://www.ncbi.nlm.nih.gov/books/NBK1269.

Picillo M, Ranieri A, Orefice G, Bonifati V, Barone P. Clinical progression of SYNJ1-related early onset atypical parkinsonism: 3-year follow up of the original Italian family. J Neurol 2014;261:823–4. doi:10.1007/s00415-014-7270-6.

Poewe W. Non-motor symptoms in Parkinson's disease. Eur J Neurol 2008;15:14–20. doi:10.1111/j.1468-1331.2008.02056.x.

Poewe W, Seppi K, Tanner CM, Halliday GM, Brundin P, Volkmann J, et al. Parkinson disease. Nat Rev Dis Primers 2017;3:1–21. doi:10.1038/nrdp.2017.13.

Polymeropoulos MH, Higgins JJ, Golbe LI, Johnson WG, Ide SE, Di Iorio G, et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. Science 1996;274:1197–9. doi:10.1126/science.274.5290.1197.

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997;276:2045–7. doi:10.1126/science.276.5321.2045.

Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30:1591–601. doi:10.1002/mds.26424.

Pringsheim T, Jette N, Frolkis A, Steeves TD. The prevalence of Parkinson's disease: A systematic review and meta-analysis. Mov Disord 2014;29:1583–90. doi:10.1002/mds.25945.

Pruitt KD, Brown GR, Hiatt SM, Thibaud-Nissen F, Astashyn A, Ermolaeva O, et al. RefSeq: an update on mammalian reference sequences. Nucleic Acids Res 2014;42:D756–63. doi:10.1093/nar/gkt1114.

Puschmann A. New Genes Causing Hereditary Parkinson's Disease or Parkinsonism. Curr Neurol Neurosci Rep 2017;17:66. doi:10.1007/s11910-017-0780-8.

Quadri M, Fang M, Picillo M, Olgiati S, Breedveld GJ, Graafland J, et al. Mutation in the SYNJ1 Gene Associated with Autosomal Recessive, Early-Onset Parkinsonism. Hum Mutat 2013;34:1208–15. doi:10.1002/humu.22373.

Quang D, Chen Y, Xie X. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. Bioinformatics 2015;31:761–3. doi:10.1093/bioinformatics/btu703.

Ramirez A, Heimbach A, Gründemann J, Stiller B, Hampshire D, Cid LP, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. Nat Genet 2006;38:1184–91. doi:10.1038/ng1884.

Reale C, Panteghini C, Carecchio M, Garavaglia B. The relevance of gene panels in movement disorders diagnosis: A lab perspective. Eur J Paediatr Neurol 2018;22:285–91. doi:10.1016/j.ejpn.2018.01.013.

Reeve A, Simcox E, Turnbull D. Ageing and Parkinson's disease: why is advancing age the biggest risk factor? Ageing Res Rev 2014;14:19–30. doi:10.1016/j.arr.2014.01.004.

Rennert H. Validation and Implementation of Whole-Exome Sequencing 2016. https://www.youtube.com/watch?v=3yKsSMf_o_8 (accessed November 8, 2019).

Riboldi GM, Di Fonzo AB. GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches. Cells 2019;8:364. doi:10.3390/cells8040364.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24. doi:10.1038/gim.2015.30.

de Rijk MC, Breteler MM, Graveland GA, Ott A, Grobbee DE, van der Meche FG, et al. Prevalence of Parkinson's disease in the elderly: The Rotterdam Study. Neurology 1995;45:2143–6. doi:10.1212/WNL.45.12.2143.

de Rijk MC, Launer LJ, Berger K, Breteler MM, Dartigues JF, Baldereschi M, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 2000;54:S21-3. doi:10.1212/WNL.54.11.21A.

Rizek P, Kumar N, Jog MS. An update on the diagnosis and treatment of Parkinson disease. CMAJ 2016;188:1157–65. doi:10.1503/cmaj.151179.

Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. Brain 2017;140:3191–203. doi:10.1093/brain/awx285.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011;29:24–6. doi:10.1038/nbt.1754.

Rocca WA. The burden of Parkinson's disease: a worldwide perspective. Lancet Neurol 2018;17:928–9. doi:10.1016/S1474-4422(18)30355-7.

Ross JP, Dupre N, Dauvilliers Y, Strong S, Ambalavanan A, Spiegelman D, et al. Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. Neurobiol Aging 2016;45:212.e13-212.e17. doi:10.1016/j.neurobiolaging.2016.04.023.

Ross JP, Dupré N, Dauvilliers Y, Strong S, Dionne-Laporte A, Dion PA, et al. RIC3 variants are not associated with Parkinson's disease in French-Canadians and French. Neurobiol Aging 2017;53:194.e9-194.e11. doi:10.1016/j.neurobiolaging.2017.01.005.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature 2011;475:348–52. doi:10.1038/nature10242.

Roze E, Coêlho-Braga MC, Gayraud D, Legrand AP, Trocello JM, Fénelon G, et al. Head tremor in Parkinson's disease. Mov Disord 2006;21:1245–8. doi:10.1002/mds.20918.

Rui Q, Ni H, Li D, Gao R, Chen G. The Role of LRRK2 in Neurodegeneration of Parkinson Disease. Curr Neuropharmacol 2018;16:1348–57. doi:10.2174/1570159x16666180222165418.

Ruiz-Martinez J, Krebs CE, Makarov V, Gorostidi A, Martí-Massó JF, Paisán-Ruiz C. GIGYF2 mutation in late-onset Parkinson's disease with cognitive impairment. J Hum Genet 2015;60:637–40. doi:10.1038/jhg.2015.69.

Ryan E, Seehra G, Sharma P, Sidransky E. GBA1-associated parkinsonism. Curr Opin Neurol 2019:1. doi:10.1097/WCO.0000000000000715.

Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, et al. Evaluation of Hybridization Capture Versus Amplicon-Based Methods for Whole-Exome Sequencing. Hum Mutat 2015;36:903–14. doi:10.1002/humu.22825.

Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 1977;74:5463–7. doi:10.1073/pnas.74.12.5463.

Santamaria R, Michelakakis H, Moraitou M, Dimitriou E, Dominissini S, Grossi S, et al. Haplotype analysis suggests a single Balkan origin for the Gaucher disease [D409H;H255Q] double mutant allele. Hum Mutat 2008;29:58–67. doi:10.1002/humu.20776.

Schapira AH, Tolosa E. Molecular and clinical prodrome of Parkinson disease: Implications for treatment. Nat Rev Neurol 2010;6:309–17. doi:10.1038/nrneurol.2010.52.

Schoenberg BS, Osuntokun BO, Adeuja AO, Bademosi O, Nottidge V, Anderson DW, et al. Comparison of the prevalence of Parkinson's disease in black populations in the rural United States and in rural Nigeria: door-to-door community studies. Neurology 1988;38:645–6. doi:10.1212/wnl.38.4.645.

Schoonen M, Seyffert AS, van der Westhuizen FH, Smuts I. A bioinformatics pipeline for rare genetic diseases in South African patients. S Afr J Sci 2019;115. doi:10.17159/sajs.2019/4876.

Schrag A, Schott JM. Epidemiological, clinical, and genetic characteristics of early-onset parkinsonism. Lancet Neurol 2006;5:355–63. doi:10.1016/S1474-4422(06)70411-2.

Schreglmann SR, Houlden H. VPS13C-Another Hint at Mitochondrial Dysfunction in Familial Parkinson's Disease. Mov Disord 2016;31:1340–1340. doi:10.1002/mds.26682.

Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 2010;7:575–6. doi:10.1038/nmeth0810-575.

Seirafi M, Kozlov G, Gehring K. Parkin structure and function. FEBS J 2015;282:2076–88. doi:10.1111/febs.13249.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29:308–11. doi:10.1093/nar/29.1.308.

Shi C, Li F, Yang J, Zhang S, Mao C, Wang H, et al. DNAJC6 mutations are not common causes of early onset Parkinson's disease in Chinese Han population. Neurosci Lett 2016;634:60–2. doi:10.1016/J.neulet.2016.09.044.

Shihab HA, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum Genomics 2014;8:11. doi:10.1186/1479-7364-8-11.

Shimura H, Hattori N, Kubo S, Mizuno Y, Asakawa S, Minoshima S, et al. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat Genet 2000;25:302–5. doi:10.1038/77060.

Shojaee S, Sina F, Banihosseini SS, Kazemi MH, Kalhor R, Shahidi GA, et al. Genome-wide linkage analysis of a Parkinsonian-pyramidal syndrome pedigree by 500 K SNP arrays. Am J Hum Genet 2008;82:1375–84. doi:10.1016/j.ajhg.2008.05.005.

Sidransky E. Gaucher disease: complexity in a “simple” disorder. Mol Genet Metab 2004;83:6–15. doi:10.1016/j.ymgme.2004.08.015.

Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. Lancet Neurol 2012;11:986–98. doi:10.1016/S1474-4422(12)70190-4.

Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson’s disease. N Engl J Med 2009;361:1651–61. doi:10.1056/NEJMoa0901281.

Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 2012;40:W452–7. doi:10.1093/nar/gks539.

Stone DL., Tayebi N, Orvisky E, Stubblefield B, Madike V, Sidransky E. Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease. Hum Mutat 2000;15:181–8. doi:10.1002/(SICI)1098-1004(200002)15:2<181::AID-HUMU7>3.0.CO;2-S.

Strauss KM, Martins LM, Plun-Favreau H, Marx FP, Kautzmann S, Berg D, et al. Loss of function mutations in the gene encoding Omi/HtrA2 in Parkinson’s disease. Hum Mol Genet 2005;14:2099–111. doi:10.1093/hmg/ddi215.

Sudhaman S, Muthane UB, Behari M, Govindappa ST, Juyal RC, Thelma BK. Evidence of mutations in RIC3 acetylcholine receptor chaperone as a novel cause of autosomal-dominant Parkinson’s disease with non-motor phenotypes. J Med Genet 2016a;53:559–66. doi:10.1136/jmedgenet-2015-103616.

Sudhaman S, Prasad K, Behari M, Muthane UB, Juyal RC, Thelma BK. Discovery of a frameshift mutation in podocalyxinlike (PODXL) gene, coding for a neural adhesion molecule, as causal for autosomal-recessive juvenile Parkinsonism. J Med Genet 2016b;53:450–6. doi:10.1136/jmedgenet-2015-103459.

Swoboda KJ, Walker MA. Neurotransmitter-Related Disorders. Swaiman’s Pediatr. Neurol. Princ. Pract. Elsevier Inc 2017;6:355–61. doi:10.1016/B978-0-323-37101-8.00044-8.

Tanner CM, Goldman SM. Epidemiology of parkinson’s disease. Neurol Clin 1996;14:317–35. doi:10.1016/S0733-8619(05)70259-0.

Tayebi N, Walker J, Stubblefield B, Orvisky E, LaMarca M., Wong K, et al. Gaucher disease with parkinsonian manifestations: does glucocerebrosidase deficiency contribute to a vulnerability to parkinsonism? Mol Genet Metab 2003;79:104–9. doi:10.1016/S1096-7192(03)00071-4.

Tekle- Haimanot R, Abebe M, Gebre-Mariam A, Forsgren L, Heijbel J, Holmgren G, et al. Community-based study of neurological disorders in rural central Ethiopia. Neuroepidemiology 1990;9:263–77. doi:10.1159/000110783.

Thaler A, Ash E, Gan-Or Z, Orr-Urtreger A, Giladi N. The LRRK2 G2019S mutation as the cause of Parkinson's disease in Ashkenazi Jews. J Neural Transm 2009;116:1473–82. doi:10.1007/s00702-009-0303-0.

Thanvi B, Lo N, Robinson T. Levodopa-induced dyskinesia in Parkinson's disease: Clinical features, pathogenesis, prevention and treatment. Postgrad Med J 2007;83:384–8. doi:10.1136/pgmj.2006.054759.

Thiruchelvam M, Richfield EK, Baggs RB, Tank AW, Cory-Slechta DA. The nigrostriatal dopaminergic system as a preferential target of repeated exposures to combined paraquat and maneb: implications for Parkinson's disease. J Neurosci 2000;20:9207–14. doi:10.1523/JNEUROSCI.20-24-09207.2000.

Trétiakoff C. Contribution a l'etude l'anatomie pathologique du locus Niger de soemmering: avec quelques déductions relatives à la pathogénie des troubles du tonus musculaire et de la maladie de Parkinson. Jouve; 1919.

Tucci A, Charlesworth G, Sheerin UM, Plagnol V, Wood NW, Hardy J. Study of the genetic variability in a Parkinson's Disease gene: EIF4G1. Neurosci Lett 2012;518:19–22. doi:10.1016/j.neulet.2012.04.033.

Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. J Neural Transm 2017;124:901–5. doi:10.1007/s00702-017-1686-y.

Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, et al. Hereditary Early-Onset Parkinson's Disease Caused by Mutations in PINK1. Science 2004;304:1158–60. doi:10.1126/science.1096284.

Valente EM, Bentivoglio AR, Dixon PH, Ferraris A, Ialongo T, Frontali M, et al. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36. Am J Hum Genet 2001;68:895–900. doi:10.1086/319522.

Vaz-Drago R, Custódio N, Carmo-Fonseca M. Deep intronic mutations and human disease. Hum Genet 2017;136:1093–111. doi:10.1007/s00439-017-1809-4.

Velayati A, Yu WH, Sidransky E. The role of glucocerebrosidase mutations in Parkinson disease and Lewy body disorders. Curr Neurol Neurosci Rep 2010a;10:190–8. doi:10.1007/s11910-010-0102-x.

Velkoff VA, Kowal PR. Aging in Sub-Saharan Africa: The Changing Demography of the Region 2006.

Vilariño-Güell C, Rajput A, Milnerwood AJ, Shah B, Szu-Tu C, Trinh J, et al. DNAJC13 mutations in Parkinson disease. Hum Mol Genet 2014;23:1794–801. doi:10.1093/hmg/ddt570.

Vilariño-Güell C, Wider C, Ross OA, Dachsel JC, Kachergus JM, Lincoln SJ, et al. VPS35 Mutations in Parkinson Disease. Am J Hum Genet 2011;89:162–7. doi:10.1016/j.ajhg.2011.06.001.

Wakabayashi K, Tanji K, Odagiri S, Miki Y, Mori F, Takahashi H. The Lewy body in Parkinson's disease and related neurodegenerative disorders. Mol Neurobiol 2013;47:495–508. doi:10.1007/s12035-012-8280-y.

Wan L, Hsu CM, Tsai CH, Lee CC, Hwu WL, Tsai FJ. Mutation analysis of Gaucher disease patients in Taiwan: High prevalence of the RecNciI and L444P mutations. Blood Cells, Mol Dis 2006;36:422–5. doi:10.1016/j.bcmd.2006.02.001.

Wang A, Costello S, Cockburn M, Zhang X, Bronstein J, Ritz B. Parkinson's disease risk from ambient exposure to pesticides. Eur J Epidemiol 2011;26:547–55. doi:10.1007/s10654-011-9574-5.

Wang Z, Liu X, Yang BZ, Gelernter J. The role and challenges of exome sequencing in studies of human diseases. Front Genet 2013;4:160. doi:10.3389/fgene.2013.00160.

Wei L, Ding L, Li H, Lin Y, Dai Y, Xu X, et al. Juvenile-onset parkinsonism with pyramidal signs due to compound heterozygous mutations in the F-Box only protein 7 gene. Parkinsonism Relat Disord 2018;47:76–9. doi:10.1016/j.parkreldis.2017.11.332.

Wider C, Skipper L, Solida A, Brown L, Farrer M, Dickson D, et al. Autosomal dominant dopa-responsive parkinsonism in a multigenerational Swiss family. Parkinsonism Relat Disord 2008;14:465–70. doi:10.1016/j.parkreldis.2007.11.013.

Wilhelmus MM, Nijland PG, Drukarch B, de Vries HE, van Horssen J. Involvement and interplay of Parkin, PINK1, and DJ1 in neurodegenerative and neuroinflammatory disorders. Free Radic Biol Med 2012;53:983–92. doi:10.1016/j.freeradbiomed.2012.05.040.

Langston JW. MPTP and parkinson's disease. Trends Neurosci 1985;8:79–83. doi:10.1016/0166-2236(85)90031-1.

Williams DR, Litvan I. Parkinsonian syndromes. Continuum (Minneap Minn) 2013;19:1189–212. doi:10.1212/01.CON.0000436152.24038.e0.

Williams DR, Watt HC, Lees AJ. Predictors of falls and fractures in bradykinetic rigid syndromes: a retrospective study. J Neurol Neurosurg Psychiatry 2006;77:468–73. doi:10.1136/jnnp.2005.074070.

Williams U, Bandmann O, Walker R. Parkinson's Disease in Sub-Saharan Africa: A Review of Epidemiology, Genetics and Access to Care. J Mov Disord 2018;11:53–64. doi:10.14802/jmd.17028.

Wilson GR, Sim JC, McLean C, Giannandrea M, Galea CA, Riseley JR, et al. Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with α-synuclein pathology. Am J Hum Genet 2014;95:729–35. doi:10.1016/j.ajhg.2014.10.015.

Winkler AS, Tütüncü E, Trendafilova A, Meindl M, Kaaya J, Schmutzhard E, et al. Parkinsonism in a population of northern Tanzania: a community-based door-to-door study in combination with a prospective hospital-based evaluation. J Neurol 2010;257:799–805. doi:10.1007/s00415-009-5420-z.

Winklhofer KF, Haass C. Mitochondrial dysfunction in Parkinson's disease. Biochim Biophys Acta 2010;1802:29–44. doi:10.1016/j.bbadis.2009.08.013.

De Witte A, Jeong K-S, Vadapalli A. OneSeq Target Enrichment Simultaneous detection of genome-wide copy number changes, cnLOH, indels, and gene mutations Application Note. 2016.

Xiong WX, Sun YM, Guan RY, Luo SS, Chen C, An Y, et al. The heterozygous A53T mutation in the alpha-synuclein gene in a Chinese Han patient with Parkinson disease: case report and literature review. J Neurol 2016;263:1984–92. doi:10.1007/s00415-016-8213-1.

Yalcin-Cakmakli G, Olgiati S, Quadri M, Breedveld GJ, Cortelli P, Bonifati V, et al. A new Turkish family with homozygous FBXO7 truncating mutation and juvenile atypical parkinsonism. Parkinsonism Relat Disord 2014;20:1248–52. doi:10.1016/j.parkreldis.2014.06.024.

Yang X, An R, Xi J, Zheng J, Chen Y, Huang H, et al. Sequencing TMEM230 in Chinese patients with sporadic or familial Parkinson's disease. Mov Disord 2017;32:800–2. doi:10.1002/mds.26996.

Yang X, Xu Y. Mutations in the ATP13A2 gene and Parkinsonism: a preliminary review. Biomed Res Int 2014;2014:371256. doi:10.1155/2014/371256.

Yonova-Doing E, Atadzhanov M, Quadri M, Kelly P, Shawa N, Musonda ST, et al. Analysis of LRRK2, SNCA, Parkin, PINK1, and DJ-1 in Zambian patients with Parkinson's disease. Park Relat Disord 2012;18:567–71. doi:10.1016/j.parkreldis.2012.02.018.

Yun SP, Kim D, Kim S, Kim S, Karuppagounder SS, Kwon SH, et al. α-Synuclein accumulation and GBA deficiency due to L444P GBA mutation contributes to MPTP-induced parkinsonism. Mol Neurodegener 2018;13:1. doi:10.1186/s13024-017-0233-5.

Zampieri S, Cattarossi S, Bembi B, Dardis A. GBA Analysis in Next-Generation Era: Pitfalls, Challenges, and Possible Solutions. J Mol Diagnostics 2017;19:733–41. doi:10.1016/j.jmoldx.2017.05.005.

Zarranz JJ, Alegre J, Gómez-Esteban JC, Lezcano E, Ros R, Ampuero I, et al. The new mutation, E46K, of α-synuclein causes parkinson and Lewy body dementia. Ann Neurol 2004;55:164–73. doi:10.1002/ana.10795.

Zhai D, Li S, Zhao Y, Lin Z. SLC6A3 is a risk factor for Parkinson's disease: A meta-analysis of sixteen years' studies. Neurosci Lett 2014;564:99–104. doi:10.1016/j.neulet.2013.10.060.

Zhang Y, Shu L, Sun Q, Zhou X, Pan H, Guo J, et al. Integrated Genetic Analysis of Racial Differences of Common GBA Variants in Parkinson's Disease: A Meta-Analysis. Front Mol Neurosci 2018;11:43. doi:10.3389/fnmol.2018.00043.

Zhang Y, Sun QY, Yu RH, Guo JF, Tang BS, Yan XX. The contribution of GIGYF2 to Parkinson's disease: a meta-analysis. Neurol Sci 2015;36:2073–9. doi:10.1007/s10072-015-2316-9.

Zhang Z-X, Dong Z-H, Román GC. Early Descriptions of Parkinson Disease in Ancient China. Arch Neurol 2006;63:782. doi:10.1001/archneur.63.5.782.

Zhou ZD, Lee JCT, Tan EK. Pathophysiological mechanisms linking F-box only protein 7 (FBXO7) and Parkinson's disease (PD). Mutat Res 2018;778:72–8. doi:10.1016/j.mrrev.2018.10.001.

Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. Am J Hum Genet 2011;89:168–75. doi:10.1016/j.ajhg.2011.06.008.

Appendices

Appendix 1: UK Parkinson's Disease Society Brain Bank criteria for the diagnosis of PD (UKPDSBB)

Step 1: Diagnosis of parkinsonian syndrome

- Bradykinesia (slowness of initiation of voluntary movement with a progressive reduction in speed and amplitude or repetitive actions)

And at least one of the following:

- Muscular rigidity
- 4–6 Hz rest tremor
- Postural instability not caused by primary visual, vestibular, cerebellar, or proprioceptive dysfunction

Step 2: Exclusion criteria for PD

- History of repeated strokes with stepwise progression of parkinsonian features
- History of repeated head injury
- History of definite encephalitis
- Oculogyric crises
- Neuroleptic treatment at onset of symptoms
- More than one affected relative*
- Sustained remission
- Strictly unilateral features after three years
- Supranuclear gaze palsy
- Cerebellar signs
- Early severe autonomic involvement
- Early severe dementia with disturbances of memory, language, and praxis
- Babinski sign
- Presence of a cerebral tumor or communicating hydrocephalus on CAT scan
- Negative response to large doses of levodopa (if malabsorption excluded)
- 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) exposure

Step 3: Supportive positive criteria of PD

Three or more required for diagnosis of definite PD:

- Unilateral onset
- Rest tremor present
- Progressive disorder
- Persistent asymmetry affecting the side of onset most
- Excellent response (70%–100%) to levodopa
- Severe levodopa-induced chorea
- Levodopa response for five years or more
- Clinical course of ten years or more

***These criteria are no longer used (Hughes et al. 2001). Adapted from Gibbs and Lees, 1988.**

Appendix 2: Ethics letter

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

Approval Letter
Progress Report

20/03/2019

Project ID: 7506

Ethics Reference #: 2002C/059

Title: Genetic analysis of inherited Parkinson's Disease and other related movement disorders

Dear Prof Jonathan Carr,

Your request for extension/annual renewal of ethics approval dated 08/03/2019 11:35 refers.

The Health Research Ethics Committee (HREC) reviewed and approved the annual progress report you submitted at a convened meeting of HREC2 on 20 March 2019.

The approval of this project is extended for a further year.

Approval date: 20 March 2019

Expiry date: 19 March 2020

Kindly be reminded to submit progress reports two (2) months before expiry date.

Where to submit any documentation

Kindly note that the HREC uses an electronic ethics review management system, *Infonetica*, to manage ethics applications and ethics review process. To submit any documentation to HREC, please click on the following link: https://applyethics.sun.ac.za.

Please remember to use your **Project ID [7506]** and Ethics Reference Number **[2002/C059]** on any documents or correspondence with the HREC concerning your research protocol.

Yours sincerely,

Mr. Francis Masiye,

HREC Coordinator,

Health Research Ethics Committee 2 (HREC2).

National Health Research Ethics Council (NHREC) Registration Number:
REC-130408-012 (HREC1) •REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review Board (IRB) Number:
IRB0005240 (HREC1) •IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects; the South African Department of Health (2006).Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition); as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).

The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.

Appendix 3: Characteristics of study participants and pedigrees

Study participants screened with the custom-designed gene panel

Lab ID number	Ethnicity	AAO	Sex	Family history	Genetic inheritance	Additional disorders	Run 1	Run 2	Positive controls	Negative controls
10.201	Caucasian	48	M	Yes[1,2]	AD	-		✔		
10.322	Mixed ancestry	48	M	Yes[1,2]	AD	-	✔			
10.334	European	46	M	Yes[1,2]	AD			✔	*SNCA* triplication	
10.334[HR]	European	46	M	Yes[1,2]	AD	-	✔		*SNCA* triplication	
10.783	European (Afrikaner)	13	M	Yes[1]	AR	Birth injury	✔			
11.861	European (Afrikaner)	63	F	Yes[1,2]	AD	Scoliosis	✔			
11.927[HR]	European	62	M	-	AD	-	✔		*LRRK2* p.G2019S	
12.726	European (Afrikaner)	43	M	Yes[1]	AD	-	✔			
12.731	Mixed ancestry	46	M	Yes[2]	AR	Epilepsy		✔		
12.799	European (Afrikaner)	49	F	Yes[1]	AD	Mood disorders- Anxiety		✔		
12.819	African	59	F	Yes[3]	AR	Hypertension				
12.951	Caucasian	50	M	Yes[1]	AD	High blood pressure, "Leaky heart" cholesterol		✔		
13.164	Caucasian	40	F	Yes[1]	AD	-		✔		

13.272	Indian	40	M	Yes[1]	AD	-		✔		
13.378	European (Afrikaner)	49	M	Yes[2]	AR	Hypercholesterolemia		✔		
13.435	European (Afrikaner)	NA	M	NA	NA	-		✔		✔
13.436	European (Afrikaner)	NA	F	NA	NA	-		✔		✔
13.437	European (Afrikaner)	NA	F	NA	NA	-		✔		✔
59.91	European	29	F	Yes[1]	AD	-	✔			
66.18	European	44	M	Yes[1,2]	AD	-	✔			
74.53	African	51	M	Yes[1]	AD	-	✔			
78.74[CHR]	African	56	F	None	AR	-	✔		*PRKN* p.G430D and *PRKN* exon 4 del	
78.84	European	50	F	Yes[1,2]	AD	-		✔		
81.58	European (Afrikaner)	38	F	Yes[2]	AD	-		✔		
82.16[CHR]	European	25	M	None	AR	-	✔		*PRKN* P113fsX163 and *PRKN* exon 3 del	
84.30	European (Afrikaner)	34	M	Yes[1,2]	AD	-		✔		
88.28	Mixed ancestry	40	M	Yes[2]	AR	-	✔			

89.01	European	47	F	Yes[1]	AD	-	✔			
90.95	Mixed ancestry	27	F	Yes[1]	AD	-	✔			
94.69	African	30	F	Yes[1]	AD	-	✔			
95.29	Caucasian	50	M	Yes[1,2]	AD	-		✔		
95.63	European (Afrikaner)	42	F	Yes[2]	AD	-		✔		

HM, Homozygous; HR, Heterozygous; CHR, Compound heterozygous; F, Female; M, Male;[1], 1st degree; [2], 2nd degree; [3],3rd degree family history; -, None; AD, Autosomal Dominant; AR, Autosomal Recessive; NA, Not applicable

Participants screened for GBA variants

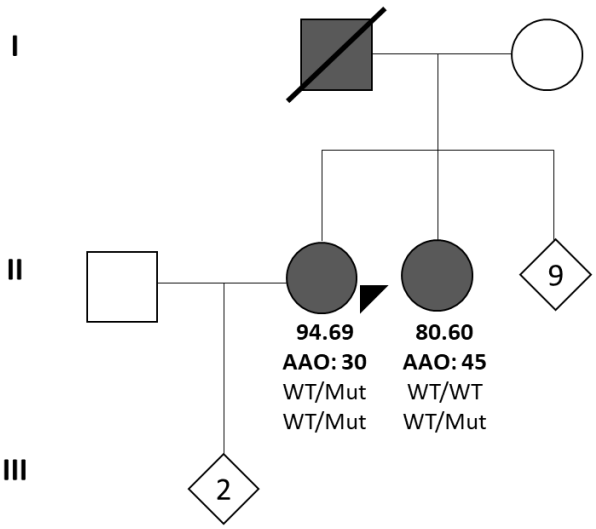
Lab ID number	Ethnicity	AAO		Sex	Family history	Genetic Inheritance	Additional disorders
10.308	African	49		F	-	NA	-
10.309	African	52		M	-	NA	-
10.310	African	45		M	-	NA	-
10.313	African	42		F	-	NA	-
10.314	African	55		F	-	NA	-
11.781	African	31		F	-	NA	-
11.830	African	57		F	-	NA	-
11.834	African	52		M	-	NA	-
11.835	African	52		F	-	NA	Hypertension
11.894	African	44		M	-	NA	-
11.895	African	49		M	-	NA	-
11.910	African	57		F	-	NA	-
11.962	African	53		M	-	NA	-
12.170	African	52		F	-	NA	Diabetes, Hypertension
12.172	African	55		F	-	NA	-
12.176	African	51		F	-	NA	-
12.177	African	45		M	-	NA	-
12.178	African	53		F	-	NA	-
12.179	African	30		F	-	NA	-
12.180	African	55		M	-	NA	-
12.486	African	35		M	-	NA	-
43.59	African	51		M	Yes[1]	AD	-
52.23	African	50		M	-	NA	-
55.52	African	42		M	-	NA	-
55.65	African	40		M	-	NA	-
60.39	African	55		M	-	NA	-
61.81	African	55		M	-	NA	-
84.52	African	57		F	-	NA	-
94.69	African	30		F	Yes[1]	AD	-
96.87	African	37		M	-	NA	-

F, Female; M, Male; [1], First degree; NA, Not applicable (Likely idiopathic PD cases: patients without family history and genetic influence is unknown);-,None

Pedigrees of the 24 study participants

1.

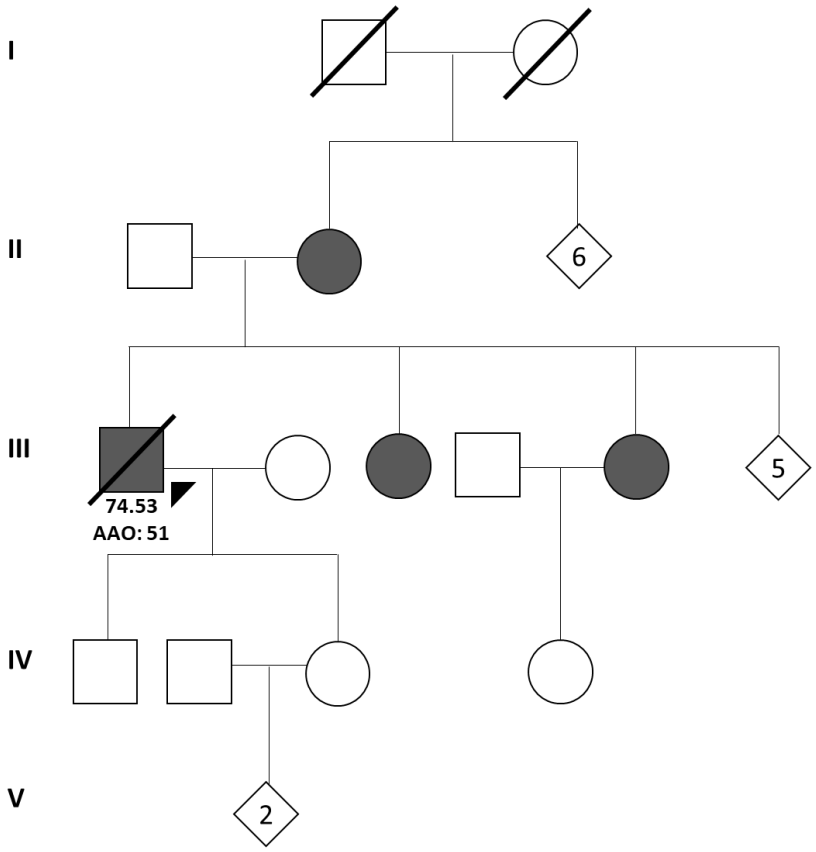
Family ZA_8



PINK1 p.P305A
GBA p.R170L (p.R131L)

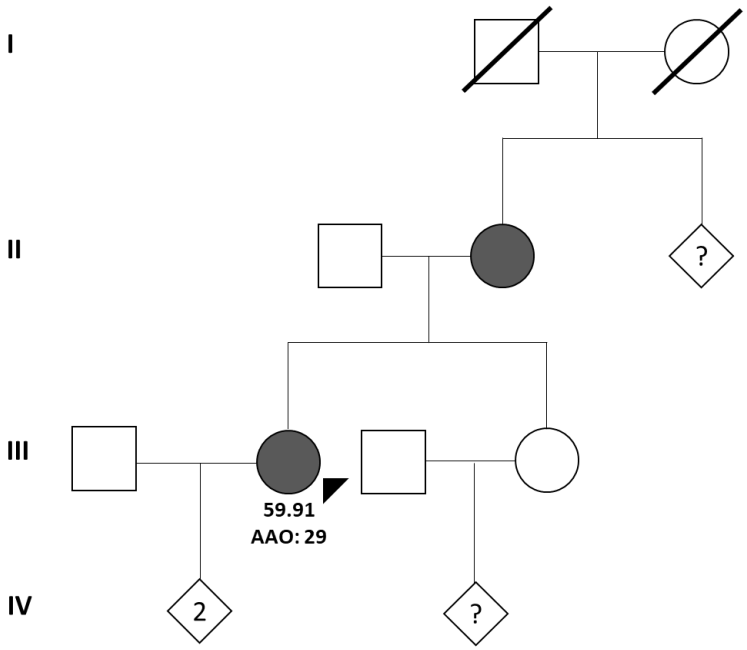
2.

Family ZA_15



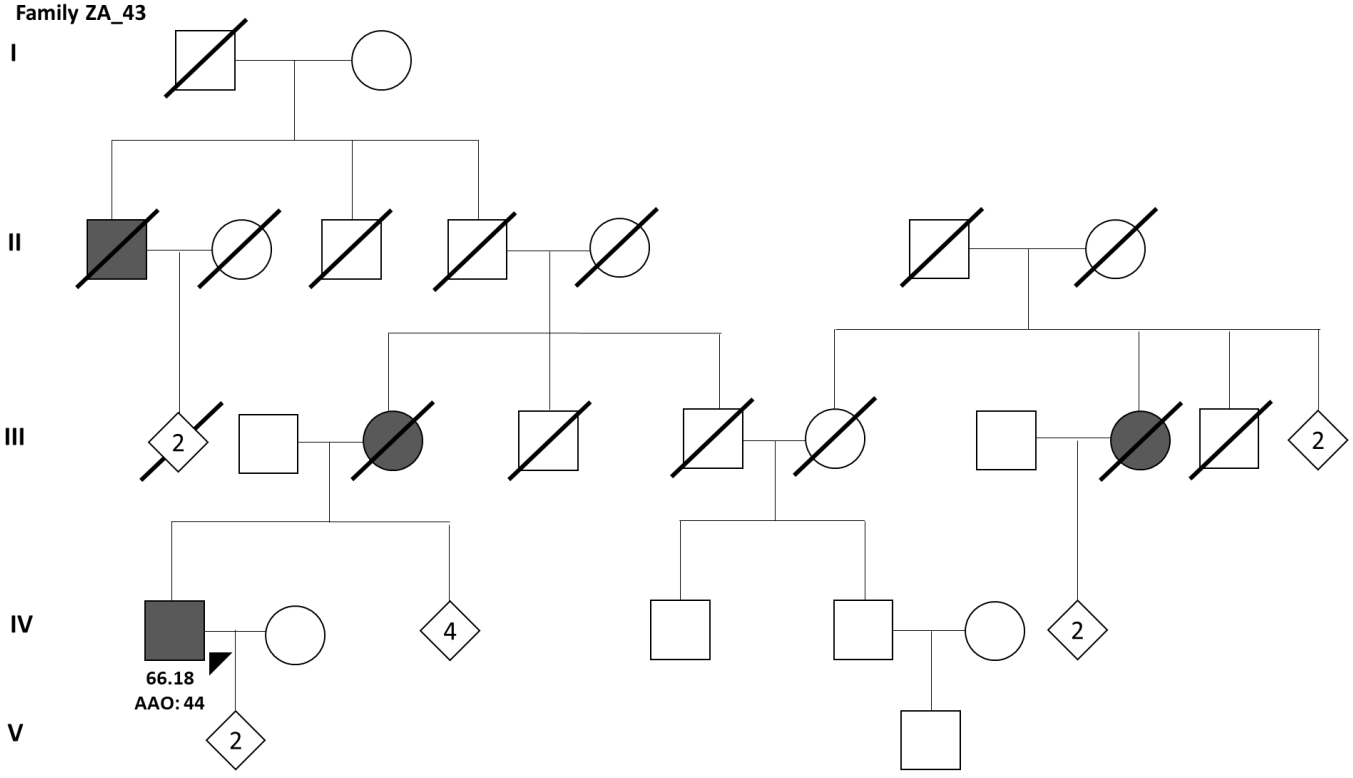
3.

Family ZA_35



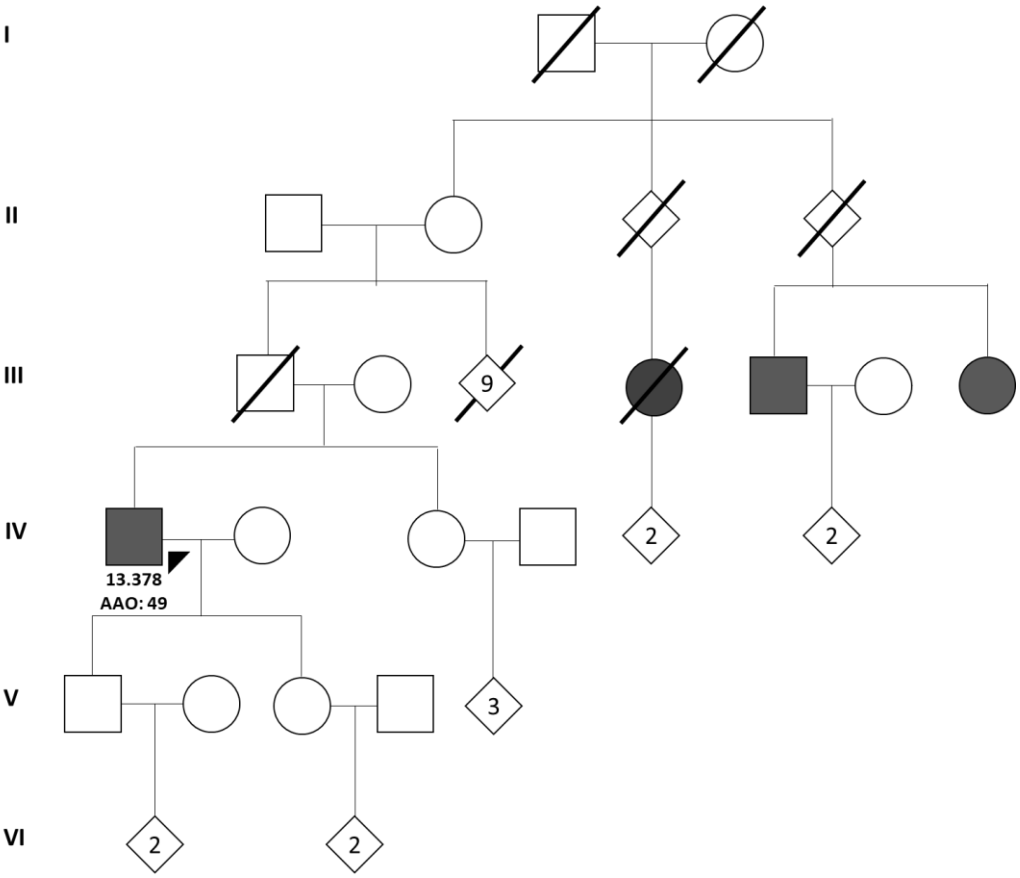
4.

Family ZA_43



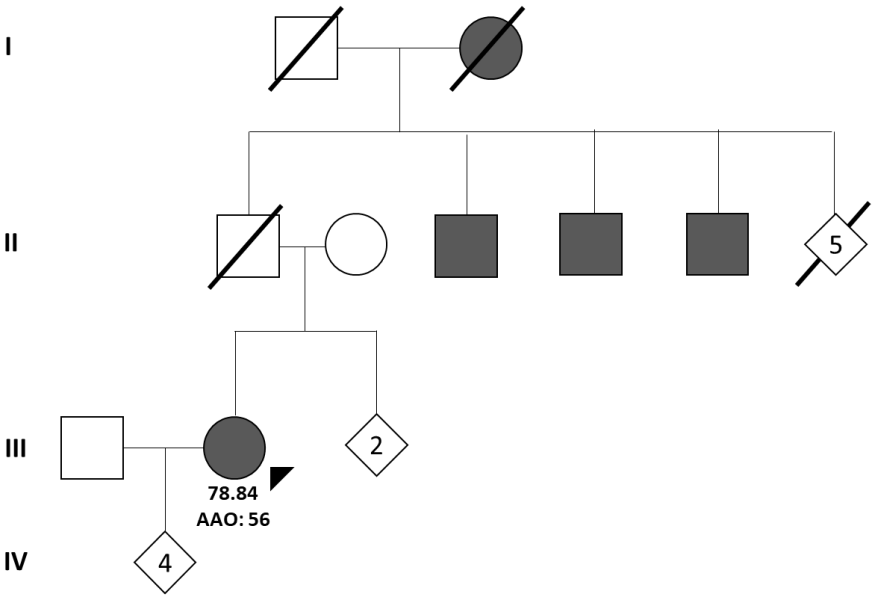
5.

Family ZA_106

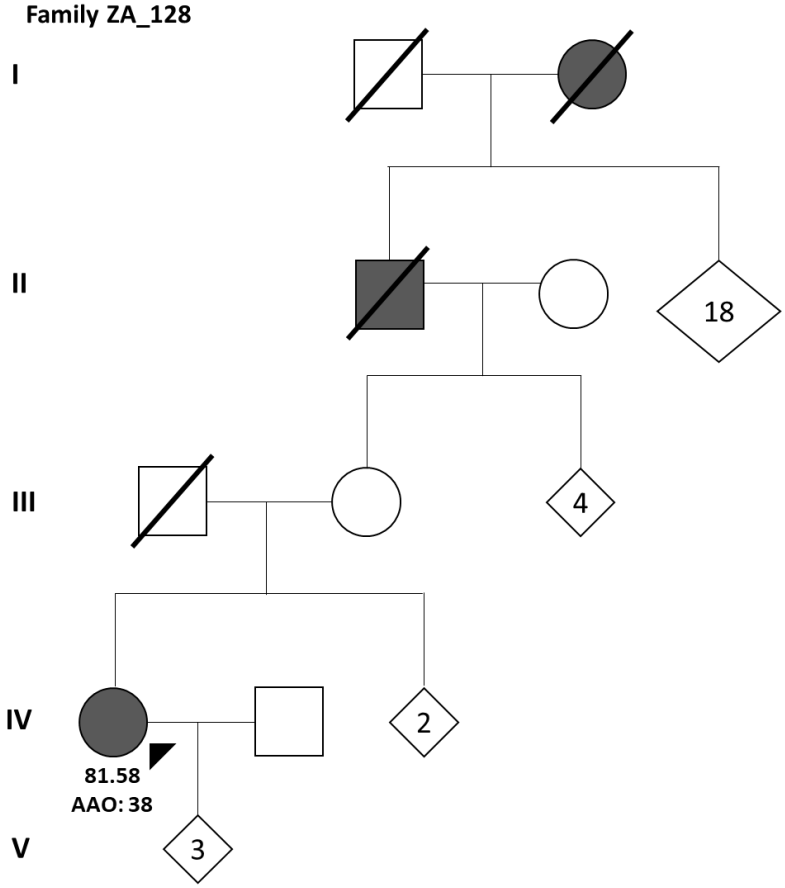


6.

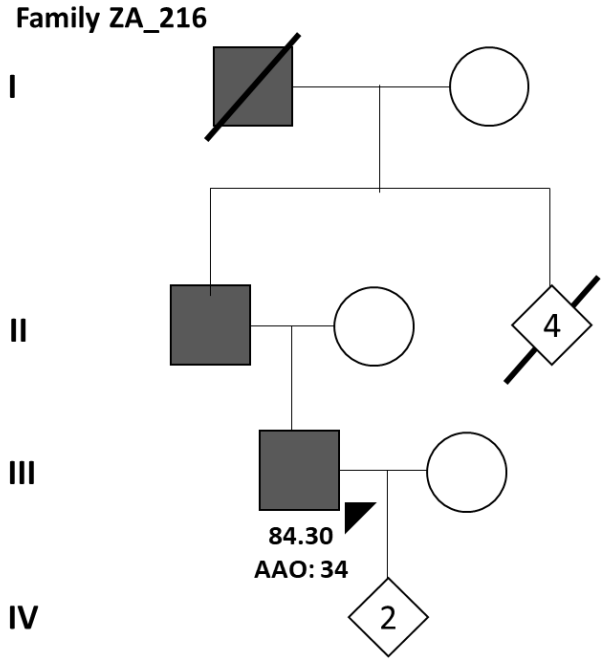
Family ZA_110



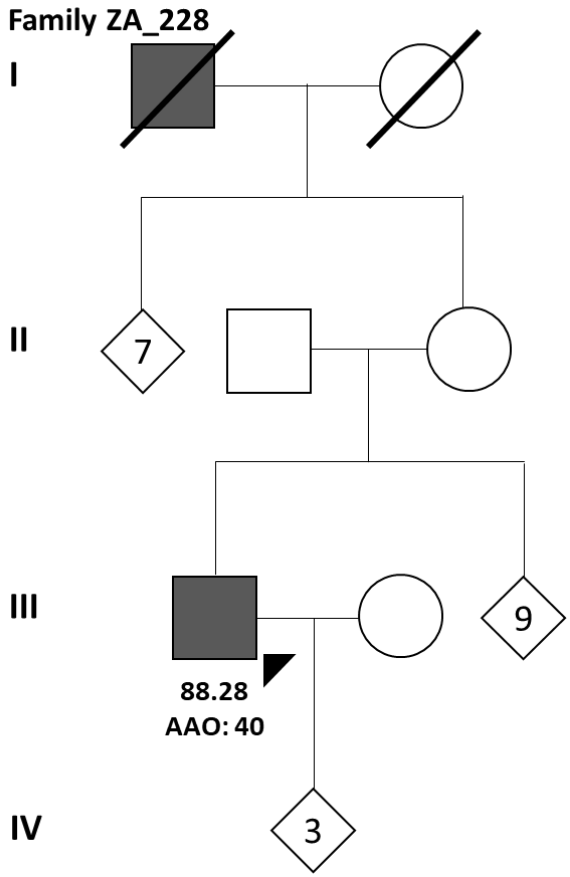
7.



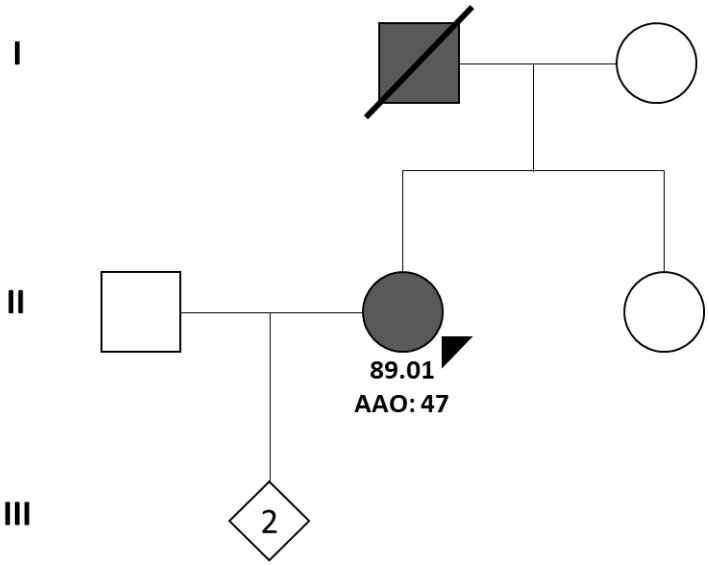
8.



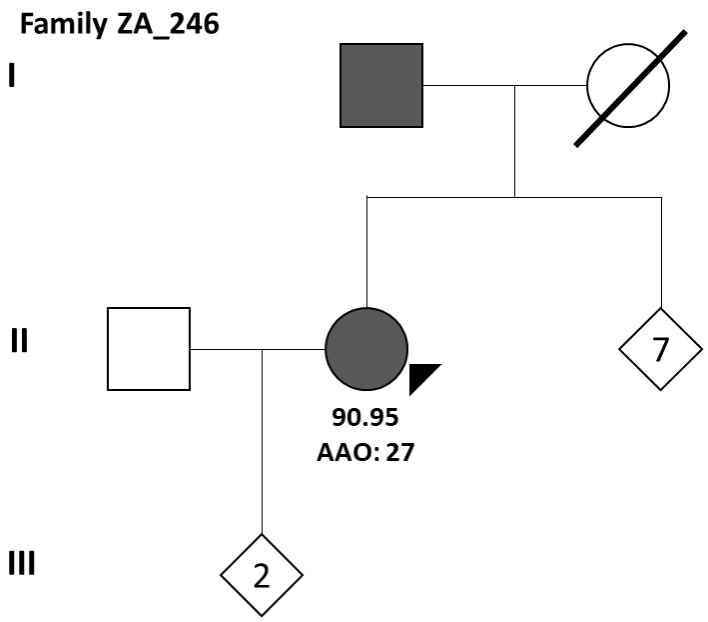
9.



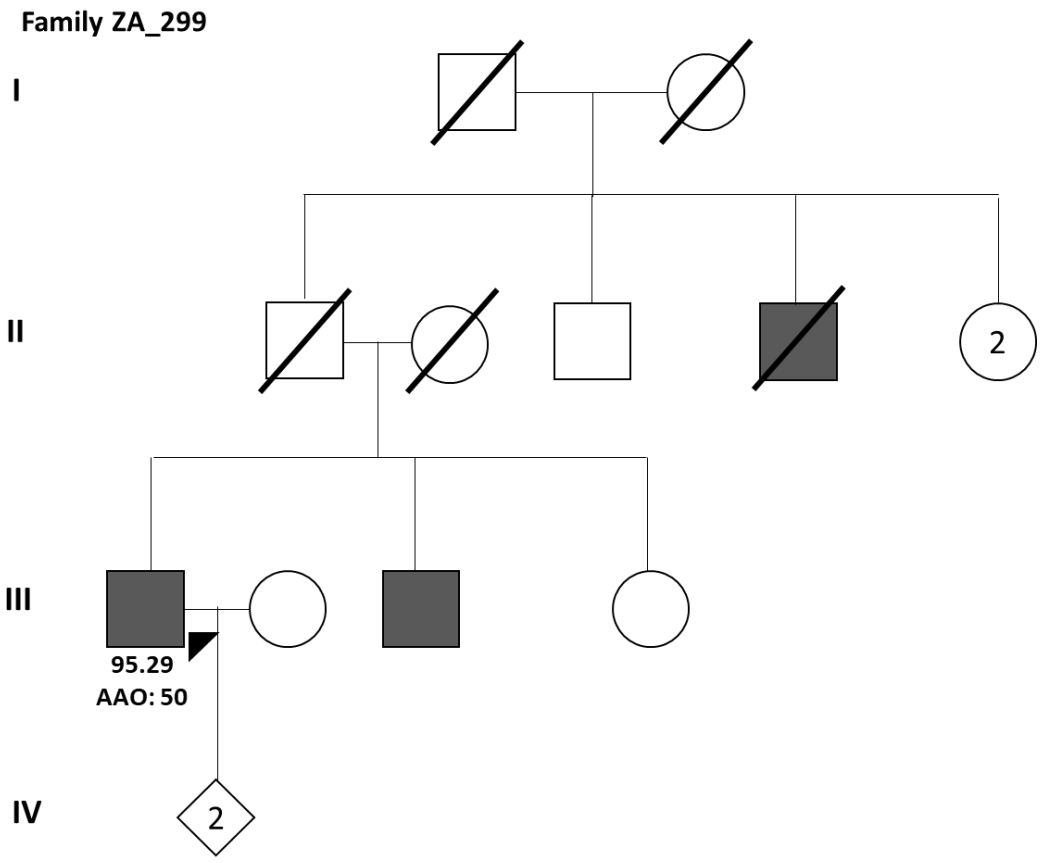
10.



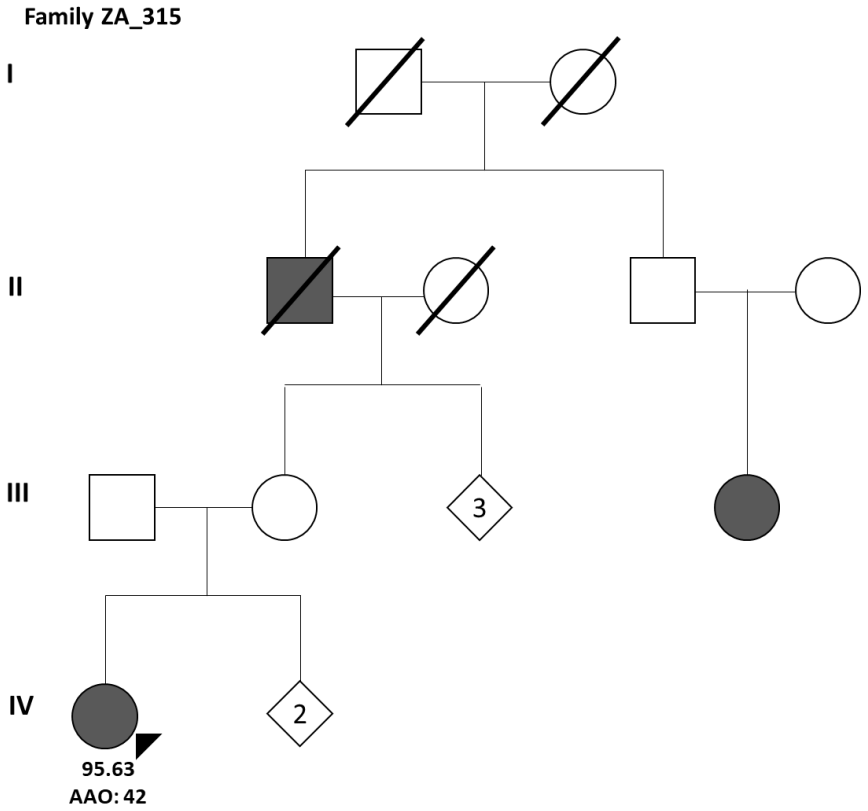
11.



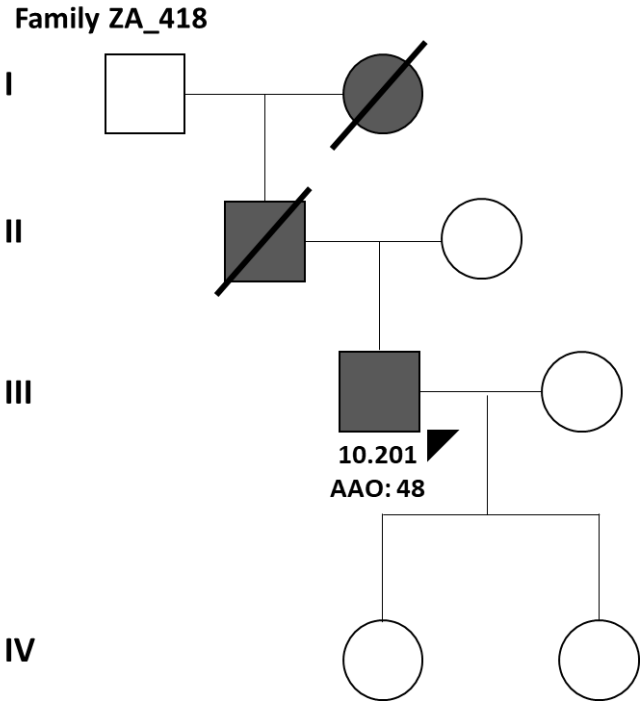
12.



13.

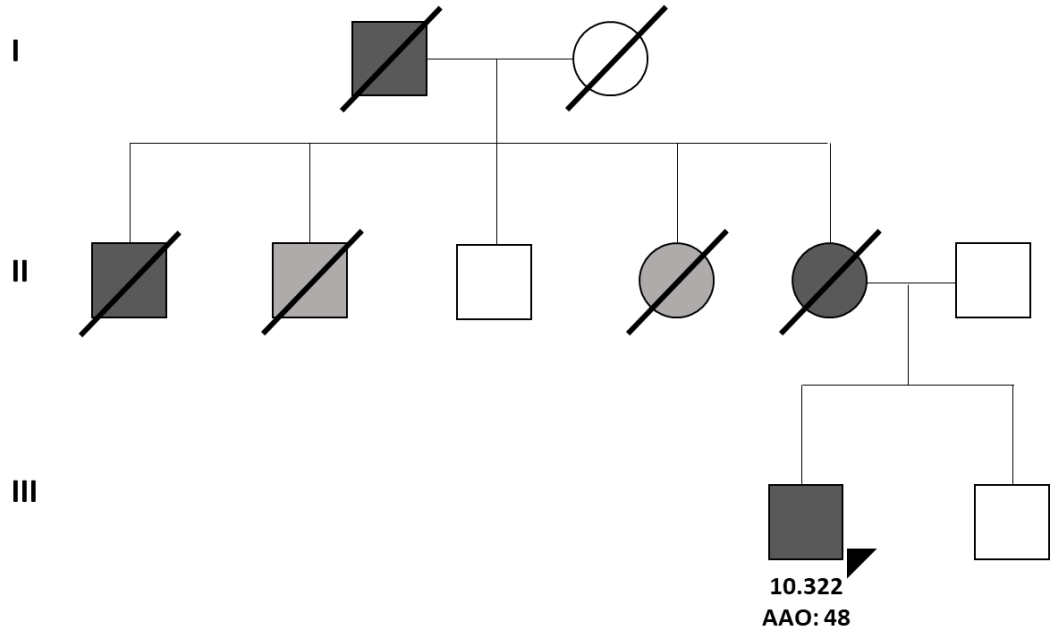


14.



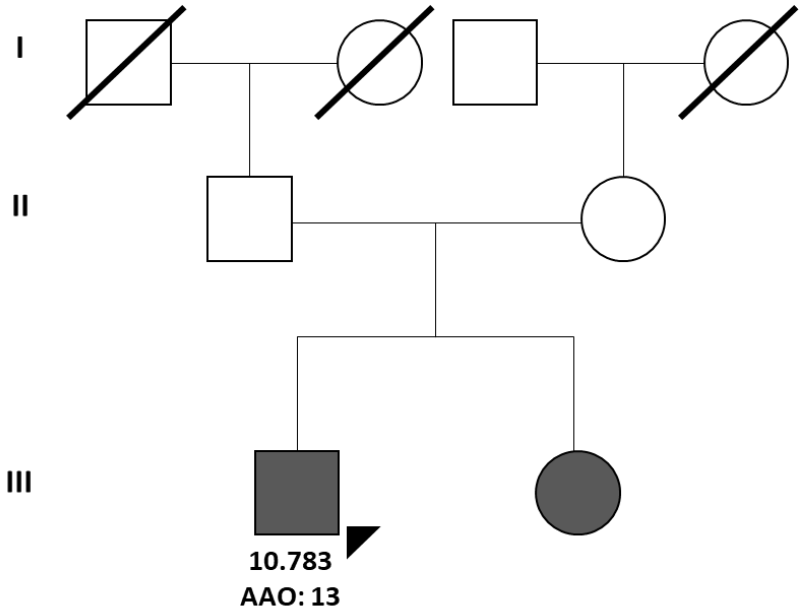
15.

Family ZA_443

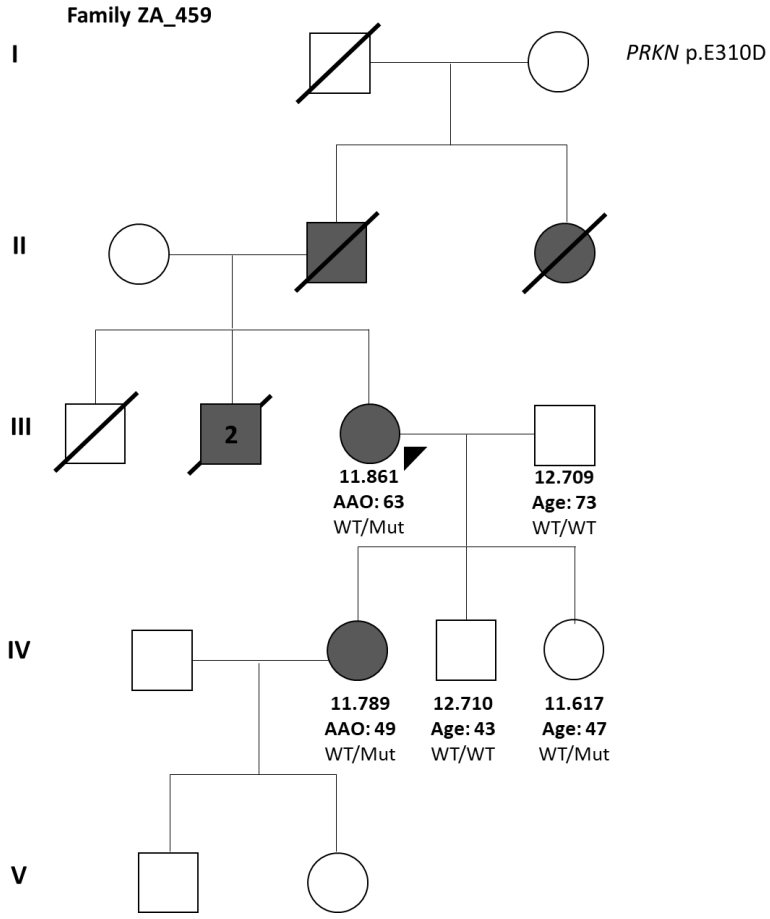


16.

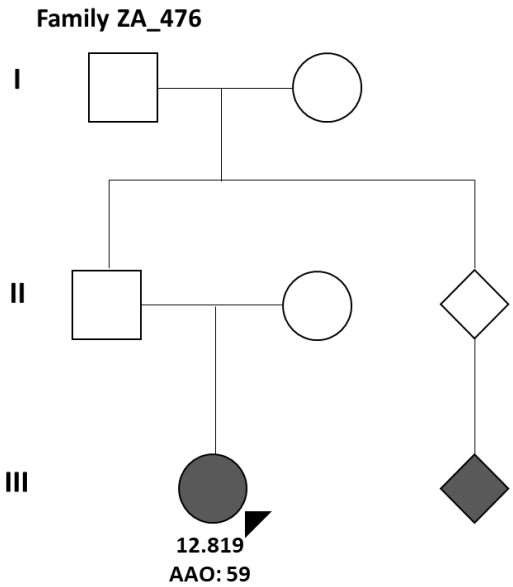
Family ZA_450



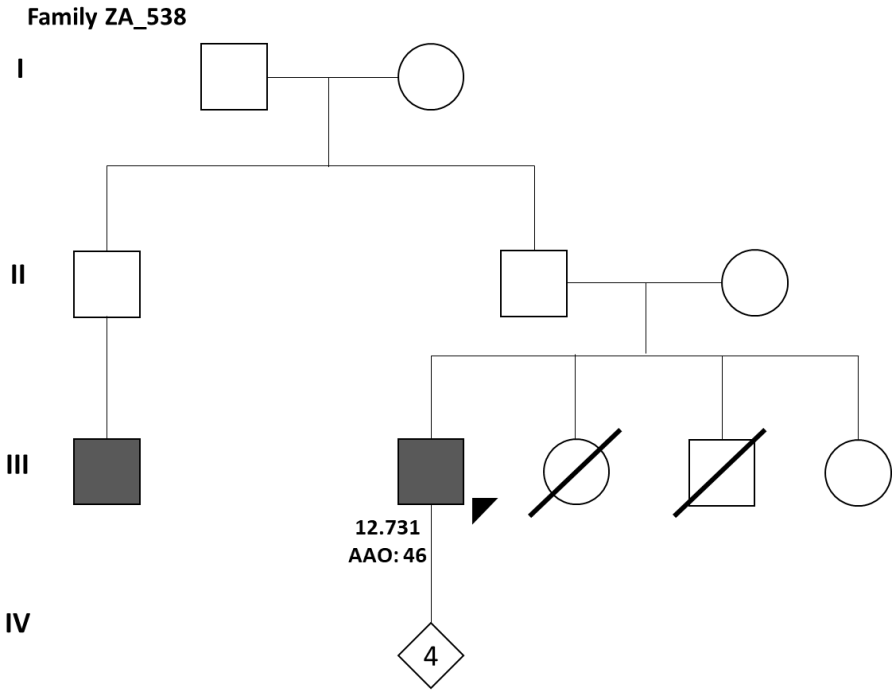
17.



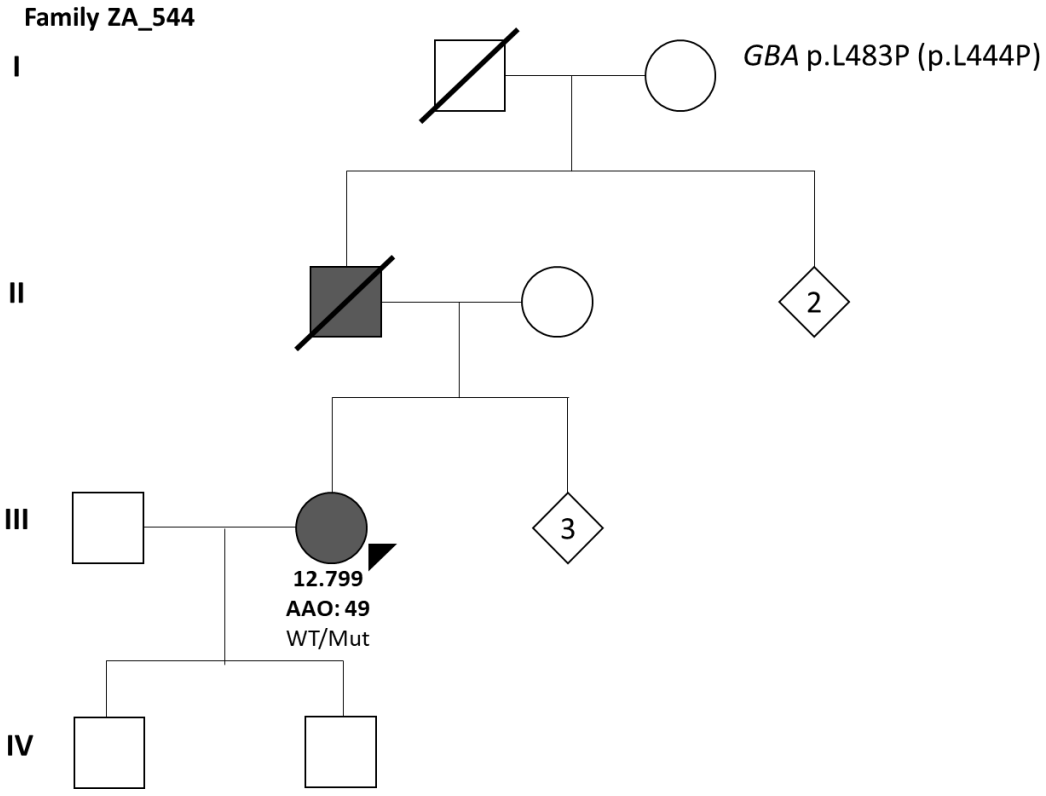
18.



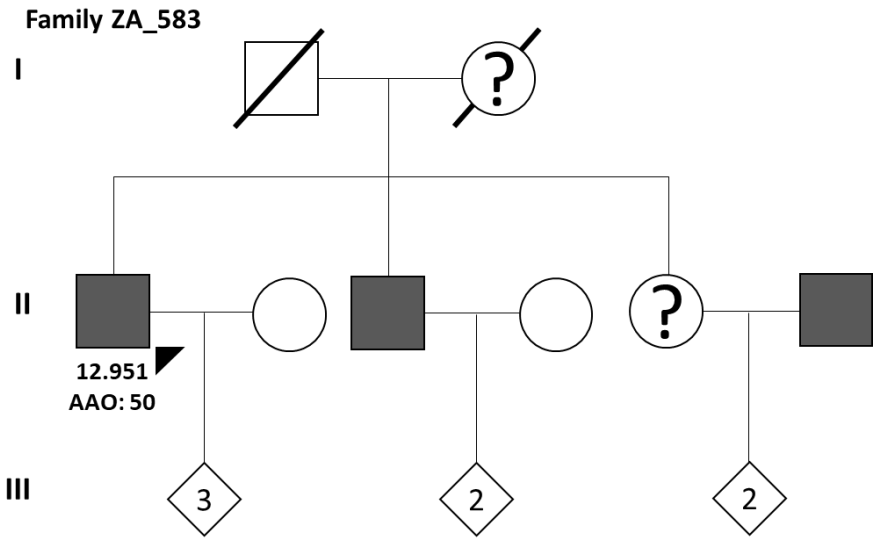
19.



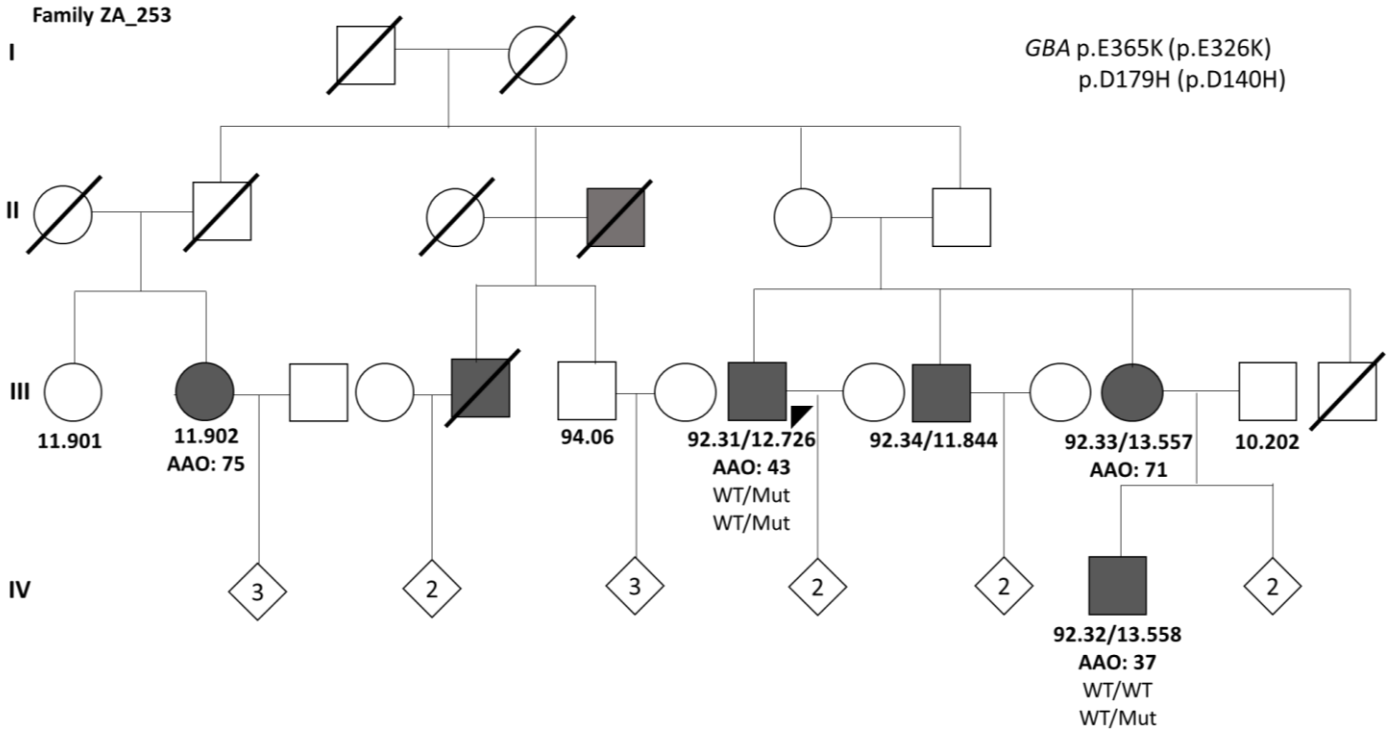
20.



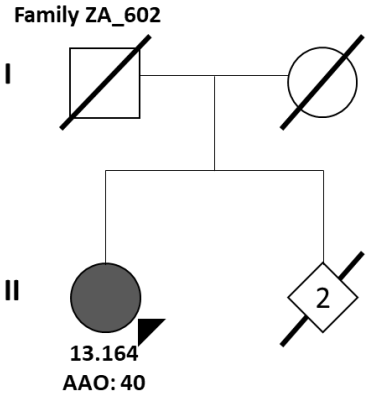
21.



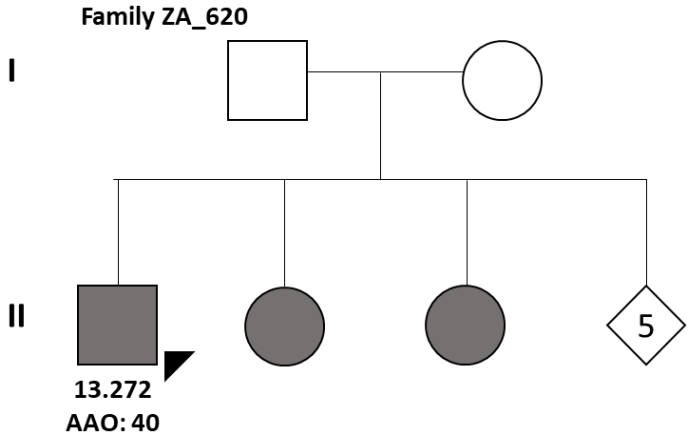
22.

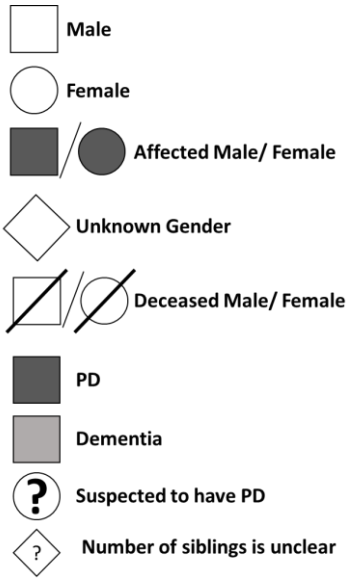


23.



24.





Appendix 4 DNA extraction methods

DNA isolation using an inhouse method- phenol/chloroform

Firstly, nuclei were isolated from whole blood. Two 5ml EDTA tubes were used to collect blood from each patient. This was then transferred into a 50 ml Falcon tube. The tube was filled up to 20 ml with ice-cold lysis buffer, gently inverted few times and incubated on ice for 5-10 minutes. This was followed by a centrifugation step at room temperature in a Beckman model TJ-6 centrifuge (Scotland, UK) at 2500-3000 rpm. Subsequently, the supernatant was disposed, and the remaining pellet was further resuspended in 20 ml of ice-cold lysis buffer. Additionally, the incubation and centrifugation steps were repeated once. Following centrifugation, the supernatant was disposed, and the pellet was resuspended in the DNA extraction buffer. At this point the DNA nuclei were either used immediately for DNA extraction or stored at -70 ˚C until the DNA was extracted when needed for genetic screening.

DNA was extracted from the previously isolated nuclei. This is achieved by adding a volume of 1000 µl of proteinase K (10 µg/ml) to the ready or thawed nuclei. The mixture is then incubated at 37˚C overnight. Subsequently, 2 ml of distilled water, 500 µl sodium-acetate (3M) and 25 µl phenol/chloroform were added to the prepared nuclei. The tube was gently mixed and inverted at 4 ˚C for 10 min on a Voss rotator (Voss of Maldon, England). At this point, an upper aqueous phase containing the DNA and an organic phase is visible in the tube. The mixture in the tube was transferred into a glass Corex to ensure that both phases are clearly distinct. Subsequently, the mixture was centrifuged for 10 minutes at 8000 rpm at 4˚C in a Sorvall RC-5B refrigerated super-speed centrifuge (rotor SS 34, Dupont Instruments). Thereafter, a sterile plastic Pasteur pipette was used to collect the upper aqueous phase containing DNA being mindful to not disturb the inorganic material. Approximately 25 ml of chloroform/octanol was then added to the aqueous phase solution. The tube containing the solution was closed with a polypropylene stopper and inverted gently for 10 minutes. Centrifugation was performed as described earlier at 4˚C and the upper aqueous phase was discarded leaving the DNA intact. Ethanol precipitation was then performed by adding two volumes of ice-cold 96% ethanol. The solution was gently inverted until a white precipitate appeared, representing the presence of DNA strands. Subsequently, removal of the DNA strands was performed using a yellow-tipped Gilson pipette. This was placed in a clean 1.5 ml Eppendorf microfuge tube. Thereafter, 1 ml of 70% ethanol was added to the DNA. This was followed by centrifugation for 3 minutes at 13 000 rpm in a Beckman microfuge and the ethanol was cautiously transferred. Additionally, 70% ethanol was added again to remove excess salts. The ethanol was removed carefully, and the DNA pellet was air-dried for 30-60 minutes at room temperature. Furthermore, the Eppendorf microfuge tube was inverted on a Carlton paper. The DNA was then resuspended in a 200 µl volume of Tris-ELDA buffer and incubated overnight for 37˚C followed by mixing

gently in a Voss rotator at 4˚C for 3 days. Subsequently, the solution was incubated at 4˚C until the DNA was completely resuspended for about 1-2 weeks. Once the DNA had fully resuspended in the buffer, its optical density (OD) was determined in a Milton Roy series 120i spectrophotometer (USA) at 260nm (OD260). Subsequently, 10 µl of DNA was diluted in 500 ul of TE buffer. The OD260 was multiplied by a factor of 2.5. Lastly, an OD260/280 of the DNA was determined to measure the purity of the DNA. An OD260/280 of about 1.8 represents a pure DNA sample.

DNA isolation using the Nucleo Spin Blood XL kit method

Briefly, to lyse the blood samples at least 10 ml of blood was collected from study participants in 50 ml Falcon tubes and equilibrated to room temperature. Subsequently, 500 ul of Proteinase K was added to each tube. Thereafter, 10 ml of BQ1 buffer was also added to the solution. At this point, the lysate solution is supposed to become a brownish colour. The solution was then vortexed vigorously for 10 seconds at 4000-5000 x g with a swing-out rotor. This was followed by a two-step incubation for 10 min at 56˚C and vertexing vigorously for 10 seconds. The samples were then incubated again for 5 min at 56 ˚C. The solution was cooled down at room temperature thereafter 10 ml ethanol (96-100%) was added. This was gently mixed by inverting the tube 10 times. For each sample 15 ml of the lysate were each added to one **NucleoSpin® Blood XL Column** kept in a collection tube. The collection was sealed with screw caps and centrifuged for 3 minutes at 3000 rpm. Subsequently, the flow-through was discarded in a blood waste container. The column was then placed into the collection tube again followed by adding 15 ml of the remaining lysate. The centrifugation was repeated and the second flow-through discarded. Once discarded the column was placed back into the collection tube. This was followed by two wash steps by adding 7.5 ml of BQ2 buffer into the column and centrifuging for 2 (for first wash) and 5 minutes (for second wash) at 3000 rpm. Subsequently, the **NucleoSpin® Blood XL Column** was dried by prolonged centrifugation for 10 minutes at 3000 rpm. The column was inserted in a new 50 ml collection tube. Thereafter, 1000 ul of preheated BE buffer (70˚C) was added into the membrane to dissolve the DNA. This was followed by incubation for 2 minutes at room temperature and centrifuged for 3000 rpm. Lastly, the DNA was transferred into labelled 1.5 Eppendorf tubes, two for each patient (A and B) and stored at 4˚C for genetic screening. The purity of the isolated DNA was determined by measuring the OD 260/280 ratio with the NanoDrop spectrophotometer.

Appendix 5: ACMG criteria guidelines for classifying variants

Criteria for Classifying Pathogenic Variants

Very strong evidence of pathogenicity	
PVS1	Null variant (nonsense, frameshift, canonical +/-1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease Caveats: • Beware of genes where LOF is not a known disease mechanism (e.g. GFAP, MYH7) • Use caution interpreting LOF variants at the extreme 3' end of a gene • Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact • Use caution in the presence of multiple transcripts
Strong evidence of pathogenicity	
PS1	Same amino acid change as a previously established pathogenic variant regardless of nucleotide change
Example:	Val->Leu caused by either G>C or G>T in the same codon
Caveat:	Beware of changes that impact splicing rather than at the amino acid/protein level
PS2	*De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, etc. can contribute to non maternity
PS3	Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered the most well-established

PS4	The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR OR does not include 1.0. See manuscript for detailed guidance. Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.
Moderate evidence of pathogenicity	
PM1	Located in a mutational hot spot and/or critical and well-established functional domain (*e.g*. active site of an enzyme) without benign variation
PM2	Absent from controls (or at extremely low frequency if recessive) (see Table 6) in Exome Sequencing Project, 1000 Genomes or ExAC Caveat: Population data for indels may be poorly called by NGS
PM3	For recessive disorders, detected in *trans* with a pathogenic variant Note: This requires testing of parents (or offspring) to determine phase
PM4	Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants
PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before Example: Arg156His is pathogenic; now you observe Arg156Cys Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level
PM6	Assumed de novo, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity	
PP1	Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease Note: May be used as stronger evidence with increasing segregation data
PP2	Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease
PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc) Caveat: As many *in-silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.
PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic etiology
PP5	Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for Classifying Benign Variants

Stand-Alone evidence of benign impact	
BA1	Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC
Strong evidence of benign impact	
BS1	Allele frequency is greater than expected for disorder (see table 6)
BS2	Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age

BS3	Well-established *in vitro* or *in vivo* functional studies shows no damaging effect on protein function or splicing
BS4	Lack of segregation in affected members of a family Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.
Supporting evidence of benign impact	
BP1	Missense variant in a gene for which primarily truncating variants are known to cause disease
BP2	Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder; or observed in *cis* with a pathogenic variant in any inheritance pattern
BP3	In-frame deletions/insertions in a repetitive region without a known function
BP4	Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc) Caveat: As many *in-silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.
BP5	Variant found in a case with an alternate molecular basis for disease
BP6	Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation
BP7	A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Rules for Combining Criteria to Classify Sequence Variants

Pathogenic

1. 1 Very Strong (PVS1) *AND*
    a. ≥1 Strong (PS1–PS4) *OR*
    b. ≥2 Moderate (PM1–PM6) *OR*
    c. 1 Moderate (PM1–PM6) and 1 Supporting (PP1–PP5) *OR*
    d. ≥2 Supporting (PP1–PP5)
2. ≥2 Strong (PS1–PS4) *OR*
3. 1 Strong (PS1–PS4) *AND*
    a. ≥3 Moderate (PM1–PM6) *OR*
    b. 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
    c. 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Likely Pathogenic

1. 1 Very Strong (PVS1) *AND* 1 Moderate (PM1–PM6) *OR*
2. 1 Strong (PS1–PS4) *AND* 1–2 Moderate (PM1–PM6) *OR*
3. 1 Strong (PS1–PS4) *AND* ≥2 Supporting (PP1–PP5) *OR*
4. ≥3 Moderate (PM1–PM6) *OR*
5. 2 Moderate (PM1–PM6) *AND* ≥2 Supporting (PP1–PP5) *OR*
6. 1 Moderate (PM1–PM6) *AND* ≥4 Supporting (PP1–PP5)

Benign

1. 1 Stand-Alone (BA1) *OR*
2. ≥2 Strong (BS1–BS4)

Likely Benign

1. 1 Strong (BS1–BS4) and 1 Supporting (BP1–BP7) *OR*
2. ≥2 Supporting (BP1–BP7)

*Variants should be classified as Uncertain Significance if other criteria are unmet or the criteria for benign and pathogenic are contradictory.

Appendix 6: PCR conditions used for amplification and primer sequences

For PINK1 and PRKN exons

Variant	exon	Forward primer sequence	Reverse primer sequence	Optimized PCR conditions	PCR Fragment Size
PINK1	4	AGGTGTTGTATCTGATGCTG	TCCCCTTGGGAGATGTATCA	94°C-5min 94°C-30sec 55°C-30sec 35 cycles 72°C-30sec 72°C-7min 4°C-hold	287 bp
PRKN	5	GGAAACATGTCTTAAGGAGT	TTCCTGGCAAACAGTGAAGA	94°C-5min 94°C-30sec 55°C-30sec 35 cycles 72°C-30sec 72°C-7min 4°C-hold	223 bp

For the amplification of GBA exons

	Forward Primer	Reverse Primer	Optimized PCR conditions	PCR Fragment Size
Large fragments (PCR 1)				
Exons 1-5	CCTAAAGTTGTCACCCATAC	AGCAGACCTACCCTACAGTTT	94°C-15min 94°C-45sec 58°C-45sec 40 cycles 72°C-3min 72°C-7min 4°C-hold	2972 bp
Exons 5-7	GACCTCAAATGATATACCTG	AGTTTGGGAGCCAGTCATTT	94°C-15min 94°C-45sec 61°C-45sec 40 cycles	2049bp

			72°C-3min 72°C-7min 4°C-hold	
Exons 8-11	TGTGTGCAAGGTCCAGGATCAG	ACCACCTAGAGGGGAAAGTG	94°C-15min 94°C-45sec 62°C-45sec 40 cycles 72°C-3min 72°C-7min 4°C-hold	1682bp
Nested PCR and sequencing primers (PCR 2)				
Exon 1	CCTAAAGTTGTCACCCATAC	CCCTCCATCTGTGCCTTGCTC	94°C-15min 94°C-45sec 55°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold	459bp
Exon 2	GAGAGTAGTTGAGGGGTGGA	CAAAGGACTATGAGGCAGAA	94°C-15min 94°C-45sec 54°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	210bp

Exon 3	ATGTGTCCATTCTCCATGTC	GGTGATCACTGACACCATTT	94°C-15min 94°C-45sec 58°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	323bp
Exon 4	GGTGTCAGTGATCACCATGG	ACGAAAAGTTTCAATGGCTCT	94°C-15min 94°C-45sec 54°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	263bp
Exon 5	GCAAGTGATAAGCAGA	AGCAGACCTACCCTACAGTTT	94°C-15min 94°C-45sec 42°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	280bp
Exon 6	CTCTGGGTGCTTCTCTCTTC	ACAGATCAGCATGGCTAAAT	94°C-15min 94°C-45sec	271bp

			52°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	
Exon 7	AGTGATCCACCTGCCTCGGC	AGTTTGGGAGCCAGTCATTT	94°C-15min 94°C-45sec 54°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	423bp
Exon 8	TGTGTGCAAGGTCCAGGATCAG	TTTGCAGGAAGGGAGACTGG	94°C-15min 94°C-45sec 55°C-45sec 30 cycles 72°C-3min 72°C-7min 4°C-hold"	294bp
Exon 9	CACAGGGCTGACCTACCCAC	GCTCCCTCGTGGTGTAGAGT	94°C-15min 94°C-45sec 59°C-45sec 30 cycles 72°C-3min	362bp

			72°C-7min 4°C-hold"	
Exon 10 **Exon 11**	GCCTCTGCAGGAGTT	 ACCACCTAGAGGGGAAAGTG	94°C-15min 94°C-45sec 43°C-45sec 35 cycles 72°C-3min 72°C-7min 4°C-hold	590bp

Appendix 7: candidate genes and panel design

Candidate PD genes

Gene Symbol	Location	Refseq IDs	Inheritance	Disease onset	Clinical Phenotype	Why we chose the PD gene	References
ATP13A2	1p36.13	NM_001141973.2	AR	Juvenile	Atypical	• Found to segregate with disease in 2 large families of Chilean and Jordanian ancestry. • A known cause of KRS of which juvenile parkinsonism is comorbidity. • Encodes an endo-lysosomal associated protein.	(Ramirez et al., 2006) (Di Fonzo et al., 2007) (Behrens et al., 2010) (Demirsoy et al., 2017)
CHCHD2	7p11.2	NM_001320327.1	AD	Early or Late	Typical	• Found in multi-incidence Japanese ADPD cases and it is indicated to be a risk factor for PD in Asian populations. • Implicated in mitochondrial apoptosis, neuronal migration, metabolism, and oxidative phosphorylation.	(Funayama et al., 2015) (Li et al., 2016) (Ikeda et al., 2017)
DJ-1	1p36.23	NM_001123377.1	AR	Juvenile or Early	Typical	• Well-established PD gene.	(Kilarski et al., 2012)
DNAJC13	3q22.1	NM_001329126.1	AD	Late	Typical	• Found to segregate with disease in a family of Dutch German-Russian ancestry with ADPD and was observed in multiple PD cases. • Implicated in the trafficking of clathrin coats on early endosomes.	(Vilariño-Güell et al., 2014) (Gustavsson et al., 2015) (Ross et al., 2016)
DNAJC6	1p31.3	NM_001256864.2	AR	Juvenile	Atypical	• Found in two ARPD affected brothers of Palestinian ancestry. • Reported in multiple early-onset PD cases. • Protein plays a role in clathrin-mediated endocytosis to integrate receptors into cells, including dopamine receptors.	(Edvardson et al., 2012) (Köroğlu et al., 2013) (Olgiati et al., 2016)

EIF4G1	3q27.1	NM_001194946.1	AD	Late	Typical	• Found to segregate with disease in a multi-incident northern French family and multiple PD cases with ADPD. • Involved in regulating mitochondrial encoding mRNA's translation.	(M. C. Chartier-Harlin et al., 2011) (Deng et al., 2015)
FBXO7	22q12.3	NM_001033024.1	AR	Juvenile	Atypical	• Well established PD gene.	(Joseph et al., 2018) (Zhou et al., 2018)
GBA	1q22	NM_000157.4	AD/AR	Early or Late	Typical	• Well established PD risk factor.	(Riboldi and Di Fonzo, 2019)
GCH1	14q22.2	NM_000161.3	AD	Early	Typical	• A common cause of DRD, PD also occurs as a neurological phenotype. • Implicated in the synthesis of monoamines, as well as dopamine production. • A novel variant was implicated in a South African family off Mixed Ancestry with DRD.	(Fabienne Clot et al., 2009) (Mencacci et al., 2014) (Guella et al., 2015) (S. Bardien et al., 2010)
GIGYF2	2q37.1	NM_001103146.1	AD	Early	Typical	• Found to segregate with disease in multiple families with ADPD. • Reported in multiple PD cases of Italian and French ethnicity. • Indicated to be a risk factor in European populations. • Involved in signalling of insulin receptors or IGFs. Insulin is implicated in the regulation of DA neurons.	(Pankratz et al., 2003) (Lautier et al., 2008) (Zhang et al., 2015) (Ruiz-Martinez et al., 2015)
HTRA2	2p13.1	NM_001321727.1	AD	Late	Typical	• Found in multiple PD cases of German ancestry displaying typical idiopathic PD. • Other variants were also found in several PD cases. • Protein indicated to interact with *PRKN* and *PINK1* to maintain mitochondrial function.	(Strauss et al., 2005) (Bogaerts et al., 2008) (Chen et al., 2014) (Lin et al., 2011) (Dagda and Chu, 2009)

LRRK2	12q12	NM_198578.4	AD	Late	Typical	• Well-established PD gene and risk factor.	(Rui et al., 2018)
PINK1	1p36	NM_032409.3	AR	Juvenile or Early	Typical	• Well-established PD gene.	(Kumar et al., 2017)
PLA2G6	22q13.1	NM_001004426.2	AR	Juvenile or Early	Atypical	• A common cause of INAD and neurodegeneration with brain accumulation of which dystonia-parkinsonism is a comorbidity. • The produced enzyme is implicated in homeostasis of cell membranes signal cell transduction.	(Morgan et al., 2006) (Paisan-Ruiz et al., 2009) (Karkheiran et al., 2015)
PRKN	6q26	NM_004562.3	AR	Juvenile or Early	Typical	• Well established PD gene.	(Dawson and Dawson, 2010)
RAB39B	Xq28	NM_171998.4	X-linked	Early	Typical or Atypical	• Known cause of X-linked parkinsonism with intellectual disability. • Implicated in the regulation of vesicular recycling pathways and the maintenance of alpha-synuclein homeostasis in the neurons.	(Puschmann, 2017)
RIC3	11p15.4	NM_001135109.3	AD	Early or late	Typical	• Found to segregate with disease in an ADPD South Indian family and in an Idiopathic early-onset PD case. • Implicated in cholinergic, glutamatergic, and dopaminergic pathways as a chaperone of a CHRNA7.	(Sudhaman et al., 2016a)
SLC6A3	5p15.33	NM_001044.5	AR	Juvenile or Early	Atypical	• Associated with multiple neuropsychiatric diseases and dopamine transporter deficiency syndrome. • Implicated to be a risk factor for PD. • Not well studied in multiple populations. • Encodes a DA transporter (DAT).	(Zhai et al., 2014) (Chang et al., 2019) (Swoboda and Walker, 2017)
SNCA	4q22.1	NM_000345.3	AD	Early or Late	Typical or Atypical	• Well established PD risk and risk factor.	(Meade et al., 2019)

SYNJ1	21q22.1	NM_001160302.1	AR	Juvenile or Early	Atypical	• Found to segregate with disease multiple ARDP families, of Sicilian, Iranian, Indian and German ancestry. • Involved in post endocytic recycling of synaptic vesicles.	(Puschmann, 2017)
TMEM230	20p13	NM_001009923.2	AD	Late	Typical	• Found to segregate with disease in a family of Dutch German-Russian ancestry where *DNAJC13* is also implicated as a causal factor. • Also, reported in multiple American and Chinese PD cases. • Involved in synaptic vesicle trafficking.	(Deng et al., 2016) (Baumann et al., 2017) (Yang et al., 2017)
VPS13C	15q22.2	NM_001160302.1	AR	Early	Atypical	• Found to segregate with disease in three unrelated families each of Turkish and French ethnicity. • Implicated in mitochondrial function.	(Lesage et al., 2016)
VPS35	16q11.2	NM_018206.5	AD	Late	Typical	• Found to segregate with disease in a large swiss family and in four additional families exhibiting ADPD. • Also, reported in multiple PD cases. • Implicated in retrograde trafficking of proteins between endosomes and trans-Golgi network.	(Vilariño-Güell et al., 2011) (Zimprich et al., 2011) (Mohan and Mellick, 2017) (Deng et al., 2013)

PD, Parkinson's disease; AD, Autosomal dominant; AR, Autosomal recessive; KRS, Kafer-Rakeb Syndrome; DRD, L-Dopa-responsive dystonia; IGF, insulin-like growth factor 1; INAD, infantile neuroaxonal dystrophy; CHRNA7, Neuronal nicotinic acetylcholine receptor subunit α-7; DAT, Dopamine transporter.

Optimization of low coverage genes

Gene ID	**Before**				**After**			
	Number of exons	Coverage	High coverage	Low coverage	Number of exons	Coverage	High coverage	Low coverage
GIGYF2	31	98.9	30	1	31	100.0	31	0
PINK1	8	99.1	8	0	8	100.0	8	0
PLA2G6	21	95.6	20	1	21	100.0	21	0
TMEM230	5	80.7	4	1	5	100.0	5	0

Before - with SureDesign default settings; After - with masking of repetitive sequences set to less stringent and probe density doubled; Coverage - bases overlapped by probes (extended +/- 100 bp) to represent likely capture; High coverage - number of regions with overlap >= 90%; Low coverage - Number of regions with overlap < 90%

Appendix 8: The rare and novel variants identified in the 24 patients

	PD genes	*ATP13A2*	*DJ-1*	*DNAJC13*	*DNAJC6*	*EIF4G1*	*FBXO7*	*GBA*	*GIGYF2*	*HTRA2*
Sample ID	**10.201**									
	10.322	p.R1067*							p.Q1210del	
	10.334									
	10.783						p.L317fs, p.L317Wfs*			
	11.861					p.S877G				
	11.927									
	12.726					p.P1065A		p.E365K (P.E326K), p.D179H (p.D140H)		
	12.731			p.R1877Q			p.R3Q			
	12.799			p.T1900M			p.R3Q	p.L483P (p.L444P)	p.Pro1225_Gln1228del p.P1219_Q1222del	
	12.819	p.P1122S, p.G158R			p.A758S			p.S251L (p.S212L), p.K13R	p.Q1210del	

								(p.K(-27)R		
	12.951								p.P1219_Q1220insPQQP	
	13.164									
	13.272								p.Q1210del	
	13.378									
	13.435									
	13.436								p.Q1210del	
	13.437									
	59.91									
	66.18								p.Q1210del	p.G22Afs*10, p.G22R, p.G22fs
	74.53							p.K13R (p.K(-27)R)	p.Q1210del	
	78.74									
	78.84								p.Q1210del	
	81.58			p.S258L		p.R402C				
	82.16									

	84.30		p.A179T							
	88.28						p.L317fs, p.L317Wfs* , p.K289R			
	89.01						p.L317fs, p.L317Wfs*		p.Q1210del	
	90.95									
	94.69		p.A56T			p.M126I	p.L317fs, p.L317Wfs*	p.R170L (p.R131L)		
	95.29								p.Q1210del	
	95.63									

	PD genes continued	***LRRK2***	***PINK1***	***PLA2G6***	***PRKN***	***RIC3***	***TMEM230***	***VPS13C***
Sample ID	**10.201**							
	10.322							p.I2746T
	10.334							
	10.783							

	11.861	p.M1646T			P.E310D			p.I1761V
	11.927							
	12.726							
	12.731	p.N2133S	p.R501Q					p.R3564H, p.P2390R, p.T2689A, p.N1453K
	12.799							
	12.819			p.V58I		p.Y218C		
	12.951		p.I610T					
	13.164							
	13.272		p.E476K			p.M134I		p.A1644V
	13.378							
	13.435							
	13.436							
	13.437							p.I2746T
	59.91							p.E1965D
	66.18							p.E1965D
	74.53	p.E899D						

	78.74			p.V58I				
	78.84							
	81.58							
	82.16							
	84.30							
	88.28							
	89.01							p.I3683V, p.R2439H
	90.95			p.V58I				
	94.69		p.P305A	p.A209T			p.S57F	
	95.29							
	95.63							p.H2337R

A, Alanine; C, Cysteine; D, Aspartic acid; E, Glutamic acid; F, Phenylalanine; G, Glycine; H, Histidine; I, Isoleucine; K, Lysine; L, Leucine; M, Methionine; N, Asparagine; P, Proline; Q, Glutamine; R, Arginine; S, Serine; T, Threonine; V, Valine; W, Tryptophan; Y, Tyrosine